

# La solution Controv3rse

Un modèle économique pour valoriser le capital culturel et humain par l'IA générative

 p.53


 p.98

 p.144

 p.190

 p.236



 Adobe Firefly

# C O N T R O V **E** R S E

3 Février 2025

<b>RESUME</b>	<b>8</b>
<b>ENJEUX</b>	<b>10</b>
1 – QUI VEUT QUE L’IA GENERATIVE REMUNERE LES AUTEURS ?	11
2 – LES TROIS VISAGES DE LA REMUNERATION DES AUTEURS	12
3 – AU-DELA DU DROIT D’AUTEUR, LE CAPITAL HUMAIN	14
<b>SOLUTION TECHNIQUE</b>	<b>16</b>
4 - SOLUTION IDEALE	17
5 – DEMONSTRATEUR	18
OBJET	18
COMMENT ÇA MARCHE	19
NOTION DE SIMILARITE	20
SIMILARITE VECTORIELLE	20
NOMBRE D’AUTEURS	21
IDENTIFICATION DES AUTEURS	21
6 – SOLUTIONS EXISTANTES	23
SYNTHESE	23
SOLUTIONS COMMERCIALISEES	24
METHODE PAY-TO-TRAIN	24
METHODE HYBRIDE SIMILARITE + PAY-TO-TRAIN	25
METHODE DE LA SIMILARITE PAR DATASET	28
METHODES EN COURS D’EXPERIMENTATION	28
VARIANTES ET EVOLUTIONS	30
7 – ACCEPTABILITE	32
8 – BREVETS	35
<b>PERSPECTIVES</b>	<b>37</b>
9 – MARCHE	38
10 – VERS LA PROPRIETE LIQUIDE	41
11 – PROMOUVOIR L’INTELLIGENCE HUMAINE	42
12 – UNE SOLUTION SACEMISANTE POUR L’EUROPE	44
<b>RECOMMANDATIONS</b>	<b>47</b>
13 – LA SOLUTION CONTROV3RSE	48
14 – ENCLENCHER LA NEGOCIATION	49
ÉTAPES	49
REMUNERATION EQUITABLE	50
TAUX EQUITABLE	50
15 – LICENCE -REMERCIEMENTS	50
<b>SUMMARY</b>	<b>54</b>
<b>CHALLENGES</b>	<b>56</b>



<b>1 – WHO WANTS GENERATIVE AI TO PAY AUTHORS?</b>	<b>57</b>
<b>2 – THE THREE FACES OF AUTHORS’ REMUNERATION</b>	<b>58</b>
<b>3 – BEYOND COPYRIGHT, HUMAN CAPITAL</b>	<b>60</b>
<b>TECHNICAL SOLUTION</b>	<b>62</b>
<b>4 - IDEAL SOLUTION</b>	<b>63</b>
<b>5 – DEMONSTRATOR</b>	<b>64</b>
OBJECT	64
HOW IT WORKS	65
CONCEPT OF SIMILARITY	66
VECTOR SIMILARITY	66
NUMBER OF AUTHORS	67
IDENTIFICATION OF AUTHORS	67
<b>6 – EXISTING SOLUTIONS</b>	<b>69</b>
SYNTHESIS	69
MARKETED SOLUTIONS	70
PAY -TO-TRAIN METHOD	70
HYBRID METHOD SIMILARITY + PAY -TO-TRAIN	71
DATASET SIMILARITY METHOD	74
METHODS CURRENTLY BEING TESTED	74
VARIANTS AND DEVELOPMENTS	76
<b>7 – ACCEPTABILITY</b>	<b>78</b>
<b>8 – PATENTS</b>	<b>81</b>
<b>PERSPECTIVES</b>	<b>83</b>
<b>9 – MARKET</b>	<b>84</b>
<b>10 – TOWARDS LIQUID PROPERTY</b>	<b>87</b>
<b>11 – PROMOTE HUMAN INTELLIGENCE</b>	<b>88</b>
<b>12 – A SACEMIZING SOLUTION FOR EUROPE</b>	<b>89</b>
<b>RECOMMENDATIONS</b>	<b>92</b>
<b>13 – THE CONTROV3RSE SOLUTION</b>	<b>93</b>
<b>14 – START THE NEGOTIATION</b>	<b>94</b>
STEPS	94
FAIR PAY	95
FAIR RATE	95
<b>15 – LICENSE – ACKNOWLEDGMENTS</b>	<b>95</b>
<b>ZUSAMMEN-FASSUNG</b>	<b>99</b>
<b>HERAUS-FORDERUNGEN</b>	<b>101</b>
<b>1 – WER MÖCHTE, DASS GENERATIVE KI AUTOREN BEZAHLT?</b>	<b>102</b>
<b>2 – DIE DREI GESICHTER DER AUTORENVERGÜTUNG</b>	<b>103</b>
<b>3 – ÜBER DAS URHEBERRECHT HINAUS: HUMANKAPITAL</b>	<b>105</b>

<b>TECHNISCHE LÖSUNG</b>	<b>107</b>
<b>4 - IDEALE LÖSUNG</b>	<b>108</b>
<b>5 – DEMONSTRATOR</b>	<b>109</b>
OBJEKT	109
SO FUNKTIONIERT ES	110
KONZEPT DER ÄHNLICHKEIT	111
VEKTORÄHNLICHKEIT	111
ANZAHL DER AUTOREN	112
IDENTIFIZIERUNG DER AUTOREN	112
<b>6 – VORHANDENE LÖSUNGEN</b>	<b>114</b>
SYNTHESE	114
VERMARKTETE LÖSUNGEN	115
PAY -TO-TRAIN -METHODE	116
HYBRIDMETHODE ÄHNLICHKEIT + PAY -TO-TRAIN	117
METHODE DER DATENSATZÄHNLICHKEIT	119
DERZEIT GETESTETE METHODEN	120
VARIANTEN UND WEITERENTWICKLUNGEN	121
<b>7 – AKZEPTANZ</b>	<b>123</b>
<b>8 – PATENTE</b>	<b>126</b>
<b>PERSPEKTIVEN</b>	<b>128</b>
<b>9 – MARKT</b>	<b>129</b>
<b>10 – AUF DEM WEG ZU LIQUIDEN VERMÖGENSWERTEN</b>	<b>132</b>
<b>11 – FÖRDERN SIE DIE MENSCHLICHE INTELLIGENZ</b>	<b>133</b>
<b>12 – EINE SACEMISIERENDE LÖSUNG FÜR EUROPA</b>	<b>135</b>
<b>EMPFEHLUNGEN</b>	<b>138</b>
<b>13 – DIE CONTROV3RSE-LÖSUNG</b>	<b>139</b>
<b>14 – BEGINNEN SIE DIE VERHANDLUNG</b>	<b>140</b>
VORGEHENSWEISE	140
FAIRE BEZAHLUNG	141
FAIRER PREIS	141
<b>15 – LIZENZ – DANKSAGUNGEN</b>	<b>141</b>
<b>RIEPILOGO</b>	<b>145</b>
<b>SFIDE</b>	<b>147</b>
<b>1 – CHI VUOLE CHE L’INTELLIGENZA ARTIFICIALE GENERATIVA PAGHI GLI AUTORI?</b>	<b>148</b>
<b>2 – LE TRE FACCE DELLA REMUNERAZIONE DEGLI AUTORI</b>	<b>149</b>
<b>3 – OLTRE IL COPYRIGHT, IL CAPITALE UMANO</b>	<b>151</b>
<b>SOLUZIONE TECNICA</b>	<b>153</b>
<b>4 - SOLUZIONE IDEALE</b>	<b>154</b>

<b>5 – DIMOSTRATORE</b>	<b>155</b>
OGGETTO	155
COME FUNZIONA	156
CONCETTO DI SOMIGLIANZA	157
SOMIGLIANZA VETTORIALE	157
NUMERO DI AUTORI	158
IDENTIFICAZIONE DEGLI AUTORI	158
<b>6 – SOLUZIONI ESISTENTI</b>	<b>160</b>
SINTESI	160
SOLUZIONI COMMERCIALIZZATE	161
METODO DI PAGAMENTO PER L'ADDESTRAMENTO	162
METODO IBRIDO SIMILARITÀ + PAY -TO-TRAIN	163
METODO DI SIMILARITÀ DEL SET DI DATI	165
METODI ATTUALMENTE IN FASE DI SPERIMENTAZIONE	166
VARIANTI E SVILUPPI	167
<b>7 – ACCETTABILITÀ</b>	<b>169</b>
<b>8 – BREVETTI</b>	<b>172</b>
<b>PROSPETTIVE</b>	<b>174</b>
<b>9 – MERCATO</b>	<b>175</b>
<b>10 – VERSO LA PROPRIETÀ LIQUIDA</b>	<b>178</b>
<b>11 – PROMUOVERE L'INTELLIGENZA UMANA</b>	<b>179</b>
<b>12 – UNA SOLUZIONE DI SACERDOTTAZIONE PER L'EUROPA</b>	<b>181</b>
<b>RACCOMANDAZIONI</b>	<b>184</b>
<b>13 – LA SOLUZIONE CONTROV3RSE</b>	<b>185</b>
<b>14 – INIZIA LA NEGOZIAZIONE</b>	<b>186</b>
PASSI	186
RETRIBUZIONE EQUA	187
TASSO EQUO	187
<b>15 – LICENZA – RICONOSCIMENTI</b>	<b>187</b>
<b>RESUMEN</b>	<b>191</b>
<b>DESAFIOS</b>	<b>193</b>
<b>1 – ¿QUIEN QUIERE QUE LA IA GENERATIVA PAGUE A LOS AUTORES?</b>	<b>194</b>
<b>2 – LAS TRES CARAS DE LA REMUNERACION DE LOS AUTORES</b>	<b>195</b>
<b>3 – MAS ALLA DE LOS DERECHOS DE AUTOR, EL CAPITAL HUMANO</b>	<b>197</b>
<b>SOLUCION TECNICA</b>	<b>199</b>
<b>4 - SOLUCION IDEAL</b>	<b>200</b>
<b>5 – DEMOSTRADOR</b>	<b>201</b>
OBJETO	201
COMO FUNCIONA	202



CONCEPTO DE SEMEJANZA	203
SIMILITUD VECTORIAL	203
NUMERO DE AUTORES	204
IDENTIFICACION DE LOS AUTORES	204
<b>6 – SOLUCIONES EXISTENTES</b>	<b>206</b>
SINTESIS	206
SOLUCIONES COMERCIALIZADAS	207
METODO DE PAGO POR FORMACION	208
METODO HIBRIDO SIMILITUD + PAGO POR CAPACITACION	209
METODO DE SIMILITUD DE CONJUNTOS DE DATOS	211
METODOS QUE SE ESTAN PROBANDO ACTUALMENTE	212
VARIANTES Y DESARROLLOS	213
<b>7 – ACEPTABILIDAD</b>	<b>215</b>
<b>8 – PATENTES</b>	<b>218</b>
<b>PERSPECTIVAS</b>	<b>220</b>
<b>9 – MERCADO</b>	<b>221</b>
<b>10 – HACIA LA PROPIEDAD LIQUIDA</b>	<b>224</b>
<b>11 – PROMOVER LA INTELIGENCIA HUMANA</b>	<b>225</b>
<b>12 – UNA SOLUCION SACEMIZADORA PARA EUROPA</b>	<b>227</b>
<b>RECOMENDACIONES</b>	<b>230</b>
<b>13 – LA SOLUCIÓN CONTROV3RSE</b>	<b>231</b>
<b>14 – INICIAR LA NEGOCIACION</b>	<b>232</b>
PASOS	232
SALARIO JUSTO	233
TARIFA JUSTA	233
<b>15 – LICENCIA – AGRADECIMIENTOS</b>	<b>233</b>
<b>概括</b>	<b>237</b>
<b>挑戰</b>	<b>239</b>
<b>1—誰希望生成式人工智慧來向作者付費？</b>	<b>240</b>
<b>2—作者報酬的三個方面</b>	<b>241</b>
<b>3—除了版權，人力資本</b>	<b>243</b>
<b>技術方案</b>	<b>244</b>
<b>4-理想的解決方案</b>	<b>245</b>
<b>5—演示者</b>	<b>246</b>
目的	246
工作原理	247
相似性的概念	248
向量相似度	248
作者人數	249

作者身份確認	249
<b>6 – 現有解決方案</b>	<b>251</b>
合成	251
已上市的解決方案	252
付費培訓方式	252
混合方法 相似度+付費訓練	253
資料集相似性方法	256
目前正在測試的方法	256
變體和發展	257
<b>7 – 可接受性</b>	<b>259</b>
<b>8. 專利</b>	<b>262</b>
<b>觀點</b>	<b>264</b>
<hr/>	
<b>9 – 市場</b>	<b>265</b>
<b>10 – 邁向液態財產</b>	<b>268</b>
<b>11. 提升人類智力</b>	<b>269</b>
<b>12 – 歐洲的薩米化解決方案</b>	<b>271</b>
<b>建議</b>	<b>274</b>
<hr/>	
<b>13 – CONTROV3RSE 解決方案</b>	<b>275</b>
<b>14 – 開始談判</b>	<b>276</b>
步驟	276
公平報酬	277
公平利率	277
<b>15 – 許可 – 致謝</b>	<b>277</b>
<b>BIBLIOGRAPHIE</b>	<b>280</b>
<hr/>	

# Résumé



Controv3rse est un think tank indépendant, formé par 70 entrepreneurs et experts de l'IA générative. Ce rapport propose une solution en réponse à la déclaration de Sam Altman, fondateur d'OpenAI, qui plaide pour un nouveau modèle économique récompensant les créateurs face à l'essor de l'IA générative.

Il faut dire que le conflit qui oppose les auteurs, dont les œuvres sont pillées, et les exploitants d'IA générative, qui ont besoin de clarté et de stabilité juridique, ne se règlera pas devant les tribunaux avant au moins 10 ans.

Dans l'intervalle, la prolifération de l'IA générative dilue l'originalité et efface les traces des œuvres sources. En parallèle, les plateformes de micro-tâches répondent à la demande massive de données au détriment des modèles basés sur la créativité et les droits d'auteur, tels que la SACEM. Le capital culturel et humain, jadis central dans la stratégie européenne, est doublement menacé d'extraction.

Or l'Économiste nobélisé Jean Tirole souligne que la richesse nationale repose sur la valorisation de l'innovation, appuyée par la propriété intellectuelle comme levier de la créativité. Pour préserver le capital culturel et humain tout en stimulant l'innovation, le nouveau modèle économique doit réinventer celui des sociétés d'auteurs.

Lloyd Shapley a introduit en 1953 un concept clé pour répartir équitablement des gains dans un jeu coopératif, basé sur la contribution de chaque joueur. La méthode dérivée, idéale pour rémunérer les auteurs de l'IA générative, a malheureusement un coût de calcul prohibitif. Des alternatives existent, comme le modèle « pay-to-train » utilisé par des banques d'images (Shutterstock, Getty). Elles rémunèrent les auteurs selon la quantité de contenu fourni, mais ignorent la qualité ou l'utilité des œuvres. Des startups et chercheurs complètent ce modèle pour offrir une répartition plus équitable fondée sur la similarité. Nous proposons au lecteur un simulateur en ligne qui lui permettra de constater par lui-même, en grandeur réelle, la faisabilité technique et économique de cette approche.

Mis en perspective, ce modèle protège les droits des auteurs et s'oppose à l'« IA ubérisante ». La « sacémisation » valorise la « propriété liquide », richesse intellectuelle de la créativité, ce qui entretient et développe le capital culturel et humain face à son extraction par les monopoles.

Nous recommandons ainsi à Sam Altman et aux exploitants d'IA générative un modèle économique en coopération avec les sociétés d'auteurs, et à toutes les parties prenantes une méthode pour enclencher les négociations dans ce nouveau cadre de réflexion.

**Enjeux**

# 1 – Qui veut que l’IA générative rémunère les auteurs ?

« Je pense que nous avons besoin d’un nouvel accord, d’une nouvelle norme, d’un nouveau protocole, peu importe comment vous voulez l’appeler, pour savoir comment les créateurs seront récompensés », a déclaré Sam Altman, fondateur d’Open AI. « Nous devons avoir de nouveaux modèles économiques qui permettront aux créateurs de générer des revenus. »<sup>1</sup>

L’idée de rémunérer les auteurs par la gen-AI est défendue par d’autres personnalités depuis plusieurs mois.

## Exploitants d’IA qui veulent rémunérer les auteurs

NAME	COMPANY	CITATION
Sam Altman	Open AI	« Je pense que nous avons besoin d’un nouvel accord, d’une nouvelle norme, d’un nouveau protocole, peu importe comment vous voulez l’appeler, pour déterminer la manière dont les créateurs seront récompensés. Nous devons avoir de nouveaux modèles économiques qui permettront aux créateurs de générer des revenus. » <sup>2</sup>
Arthur Mensch	Mistral AI	« Nous avons intérêt à avoir accès à des contenus de qualité. Nous avons donc entamé des discussions avec des fournisseurs de contenus, dans l’édition et dans la presse. Il y a des synergies et des mécanismes de partage de valeur à trouver. » <sup>3</sup>
David Holz	Midjourney	« Il n’existe pas vraiment de moyen de récupérer une centaine de millions d’images et de savoir d’où elles viennent. Ce serait bien si les images comportaient des métadonnées intégrées sur le propriétaire des droits d’auteur ou autre. Mais ce n’est pas possible, il n’existe pas de registre. Il n’existe aucun moyen de trouver une image sur Internet, puis de la retracer automatiquement jusqu’à son propriétaire et de pouvoir ensuite faire quoi que ce soit pour l’authentifier. » <sup>4</sup>
Chris Dixon	a16z	« Le mouvement [du Web 3] a le potentiel de ramener l’esprit des débuts de l’Internet, de garantir les droits de propriété des créateurs, de récupérer la propriété et le contrôle des utilisateurs et de briser l’emprise des grandes entreprises centralisées sur nos vies. » <sup>5</sup> .
Neal Stephenson	Lamina1 (inventor of metaverse)	« Vous pouvez en fait suivre le degré d’influence qu’une image ou une entrée donnée a eu sur le résultat final. Et puis, si ce résultat final vaut quelque chose, alors peut-être qu’il y a un moyen de donner du crédit et une compensation proportionnelle à ces entrées [...] Disons que j’écrivais un livre qui contenait une épée magique qui était juste décrite dans le livre. [...] Quelqu’un pourrait alors créer un élément qui pourrait être vendu sur la boutique d’éléments Unreal [...] Vous avez des concepteurs sonores qui doivent faire quelque chose de similaire avec les sons qu’elle fait quand elle est utilisée. Et vous avez des programmeurs qui doivent, à l’aide de plans ou de C++, intégrer l’épée dans le jeu pour qu’elle soit réellement capable de faire quelque chose et de contribuer à l’expérience. Ainsi, à chaque étape, davantage de valeur est ajoutée. Et à la fin de ce processus, vous avez quelque chose qui pourrait réellement générer des revenus. Et lorsque ces revenus apparaissent enfin, ce que vous aimeriez qu’ils fassent, c’est se propager en arrière. Et vous aimeriez que les différentes personnes qui ont contribué à la chaîne de valeur soient rémunérées d’une manière ou d’une autre » <sup>6</sup>



NAME	COMPANY	CITATION
Jaron Lanier	Microsoft	« Les gens seront payés pour leurs données et paieront pour des services qui nécessitent des données provenant d'autres personnes. L'attention des individus sera guidée par leurs intérêts auto-définis plutôt que par des plateformes manipulatrices redevables à des annonceurs ou à d'autres tiers. Les plateformes recevront des données de meilleure qualité avec lesquelles former leurs systèmes d'apprentissage automatique. » Lanier reconnaît que même les chercheurs en matière de dignité des données ne parviennent pas à s'entendre sur la manière de démêler tout ce que les modèles d'IA ont absorbé ou sur le degré de détail nécessaire à en rendre compte. Pourtant, Lanier pense que cela pourrait être fait – progressivement. « Nous devons calculer et présenter la provenance des sources humaines qui ont été les plus importantes pour un résultat d'IA donné. Nous ne le faisons pas actuellement. Nous pouvons le faire de manière efficace et efficiente, c'est juste que nous ne le faisons pas. Il doit s'agir d'une décision sociétale de passer à cette étape » <sup>7</sup> .
Yat Siu	Animoca Brands	« Les données sont les ressources les plus précieuses. Pour commencer, elles alimentent toute l'intelligence artificielle dont nous entendons tant parler. Sans données, pas de ChatGPT . Les données sont le nouveau travail. Et nous ne sommes pas rémunérés équitablement pour cela. [...] Il existe une corrélation évidente entre les droits de propriété et la richesse des nations. » « Imaginez que vous alliez sur Facebook et qu'à la fin de la journée, vous réalisiez que vous venez de gagner mille dollars pour l'entreprise. Votre relation avec Facebook changerait complètement. Vous exigeriez probablement votre juste part. » <sup>8</sup>
Yann Le Cun	Meta	« un modèle de propagation des droits [d'auteur] dans un sens et de revenus dans l'autre » <sup>9</sup>
Trip Adler	former CEO of Scribd	« La quatrième loi est un ensemble de principes directeurs sur la manière dont les entreprises d'IA peuvent utiliser et s'entraîner sur le contenu créé par l'homme. La quatrième loi, inspirée des trois lois des robots de l'auteur de science-fiction Isaac Asimov, stipule que les humains devraient avoir le droit de consentir et de contrôler la manière dont l'IA utilise leur travail, d'être rémunérés (si demandés) et crédités pour leur travail (si un livre est référencé dans le résultat, il devrait y avoir un lien pour l'acheter). » <sup>10</sup>

Compte-tenu du leadership de Sam Altman, l'idée de rémunérer les auteurs prend davantage de consistance, bien qu'il considère que cette rémunération ne serait due que lorsque le style de l'auteur est plagié<sup>11</sup>, ce qui est très limitatif.

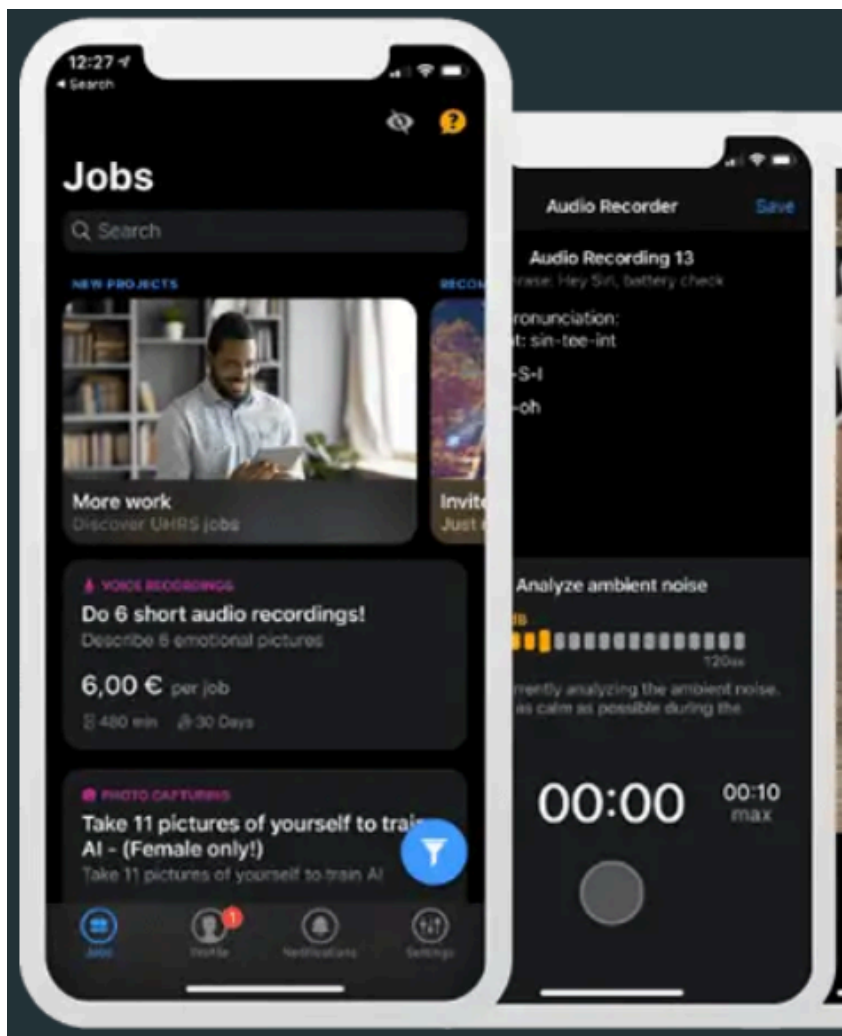
Il s'interroge néanmoins sur le modèle économique de sa mise en oeuvre. L'objet de cette note est de proposer ce nouveau modèle économique dans le cadre de l'AI Act<sup>12</sup>.

## 2 – Les trois visages de la rémunération des auteurs

S'il est traité sur le seul plan juridique, le conflit autour du scraping, qui oppose auteurs et exploitants d'IA générative<sup>13</sup>, durera au moins dix ans.

Dans l'intervalle, la filière va se segmenter entre créateurs de données, plateformes, agrégateurs, négociateurs, entraîneurs, affineurs, exploitants et intégrateurs d'IA générative, ce qui va diluer l'acte de création. Le patrimoine culturel aura été mélangé, diffusé et recyclé dans des contenus dont les traces originelles auront été effacées.

Par crainte d'une dégénérescence des contenus d'apprentissage<sup>14</sup>, une nouvelle filière de micro-tâches se développe déjà pour la production massive de contenus.



©Clickworker 2024

Les plateformes de la gig economy pour l'IA sont fondées sur une logique de commande. Clickworker<sup>15</sup> paye directement les utilisateurs 6€ pour décrire 6 images à haute voix et en poster l'enregistrement. Ou pour se prendre en photo en train de faire des exercices physiques. Gigwalk et Premise rémunèrent 4\$ pour prendre la photo d'un embouteillage à une intersection<sup>16</sup>. Le résultat sert à entraîner les IA.

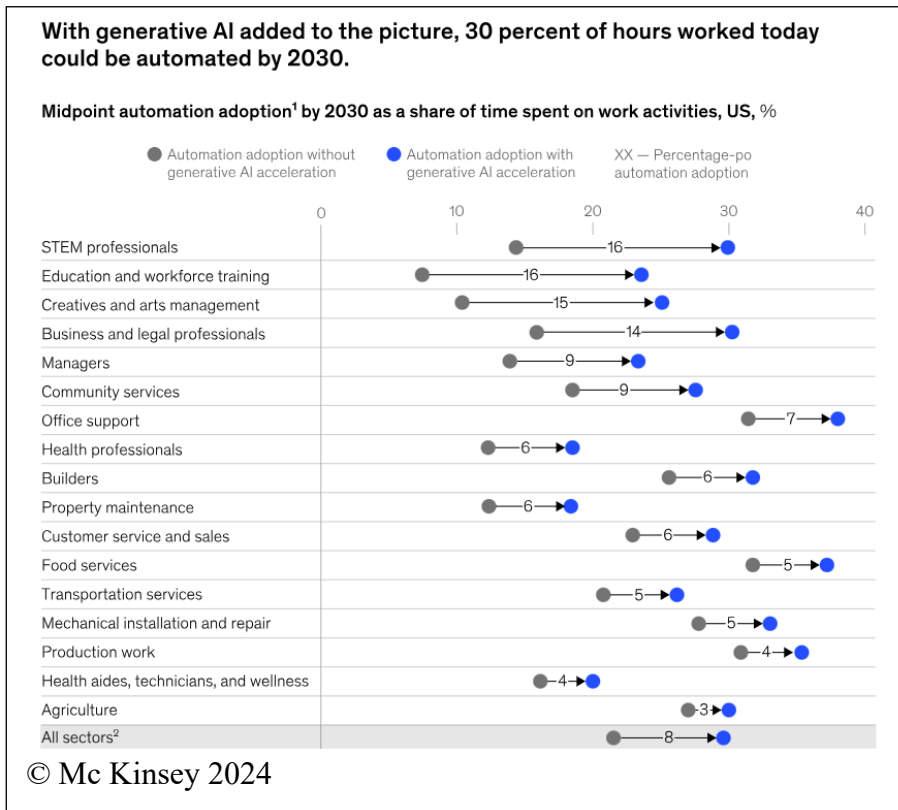
Ce modèle de type « Uber des données » est en pleine croissance grâce au besoin massif de données des exploitants d'IA générative.

C'est l'inverse des sociétés d'auteur comme la SACEM, fondés sur l'offre et la propriété intellectuelle des auteurs, qui tracent cette propriété, collectent les redevances, et rémunèrent les ayants-droits selon des grilles de répartition prédéfinies.

Le premier modèle favorise la demande, l'exécution, la marchandisation immédiate, le néo-taylorisme cognitif, la normalisation du travail. Le second favorise l'offre, l'initiative, la propriété intellectuelle, la créativité, la diversité culturelle. Le premier exploite le capital culturel, le second le cultive.

Parmi d'autres modèles économiques (voir plus bas), les scraping, Uber des données et Sociétés d'auteurs sont les trois visages principaux de la rémunération – ou non – des auteurs, dont le déséquilibre aggrave les tensions actuelles.

### 3 – Au-delà du droit d’auteur, le capital humain



Face au développement de l’IA générative, les auteurs et la créativité sont en première ligne. Suivront l’artisanat, les métiers d’art, l’expertise, la recherche et tous les métiers ayant un savoir-faire<sup>17</sup>, dans toutes les filières économiques.

En se généralisant, le modèle de l’« Uber des données » donnera naissance à une forme d’IA ubérisante, de la même manière que la société Uber a initié jadis le phénomène d’ubérisation.

Personnaliser par exemple l’IA générative d’entreprise avec toutes ses archives de données permet d’obtenir un brouillon pour chaque nouveau message, compte-rendu, projet, présentation, tarification<sup>18</sup>. Les

rectifications des professionnels entraînent l’IA jusqu’à ce que le taux d’erreur devienne acceptable. Les gains de productivité sont sensibles mais, en contrepartie, les positions de l’entreprise sur les chaînons stratégiques de la création de valeur diminuent au bénéfice de l’exploitant de l’IA.

Comme tous ces savoir-faire sont le croisement de savoirs, de pratiques, de matériaux d’organisation sociale, c’est le tissu des entreprises qui est en jeu<sup>19</sup>. C’est comme si, au moment de la révolution industrielle, un cartel de propriétaires des machines à vapeur contrôlait l’organisation tayloriste du travail. C’est ce qu’anticipe la bourse pour les Big Tech. En accordant des capitalisations en milliers de milliards, elle ne valorise plus des entreprises, mais des monopoles.

L’idée que les machines ont le « droit d’apprendre » gratuitement<sup>20</sup> va se distiller. Dans dix ans, même si les auteurs finissent par emporter leur bataille judiciaire face au scraping, leur compensation se fera au regard des comparables fournis par les Uber des données, c’est-à-dire peu de chose. On se rendra compte rétrospectivement que le droit d’auteur existant n’aura été qu’une ligne Maginot facile à contourner.

L’économie dispose d’un capital culturel et humain<sup>21</sup>, identifié depuis longtemps comme un actif considérable, au point d’avoir été placé au centre de la stratégie économique européenne<sup>22</sup>. Cette stratégie, trop précoce, a finalement été oubliée faute de modèle économique. Sous-valorisé et mal défendu, ce capital est maintenant menacé d’extraction par l’IA ubérisante. Pour chaque entreprise, pays ou continent, refuser la technologie c’est se condamner à un décrochage durable, comme celui de la Chine pendant deux siècles. L’accepter sans contrepartie, c’est se laisser coloniser sans combattre.

Le prix Nobel Jean Tirole nous enseigne que 1/ la richesse des nations dépend de leur capacité à capter la valeur au niveau de l’innovation<sup>23</sup>. 2/ l’innovation provient de l’initiative individuelle, car le propre du



travail créatif est que l'on ne sait pas ce qu'on va trouver<sup>24</sup>. 3/ la propriété intellectuelle est un mal nécessaire pour stimuler cette créativité<sup>25</sup>. 4/ empêcher une entreprise de commercialiser ses produits tant qu'elle n'a pas payé ses redevances est une arme très efficace contre les Gafam<sup>26</sup>. 5/ des pools de propriété intellectuelle avec autorisation de licences individuelles leur imposeraient d'acquérir des licences<sup>27</sup>, mais à un prix concurrentiel<sup>28</sup>.

Les nations qui veulent valoriser leur capital humain et culturel doivent trouver une solution sacémisante moderne, dont le modèle économique concurrencera efficacement l'IA ubérisante.

# **Solution technique**

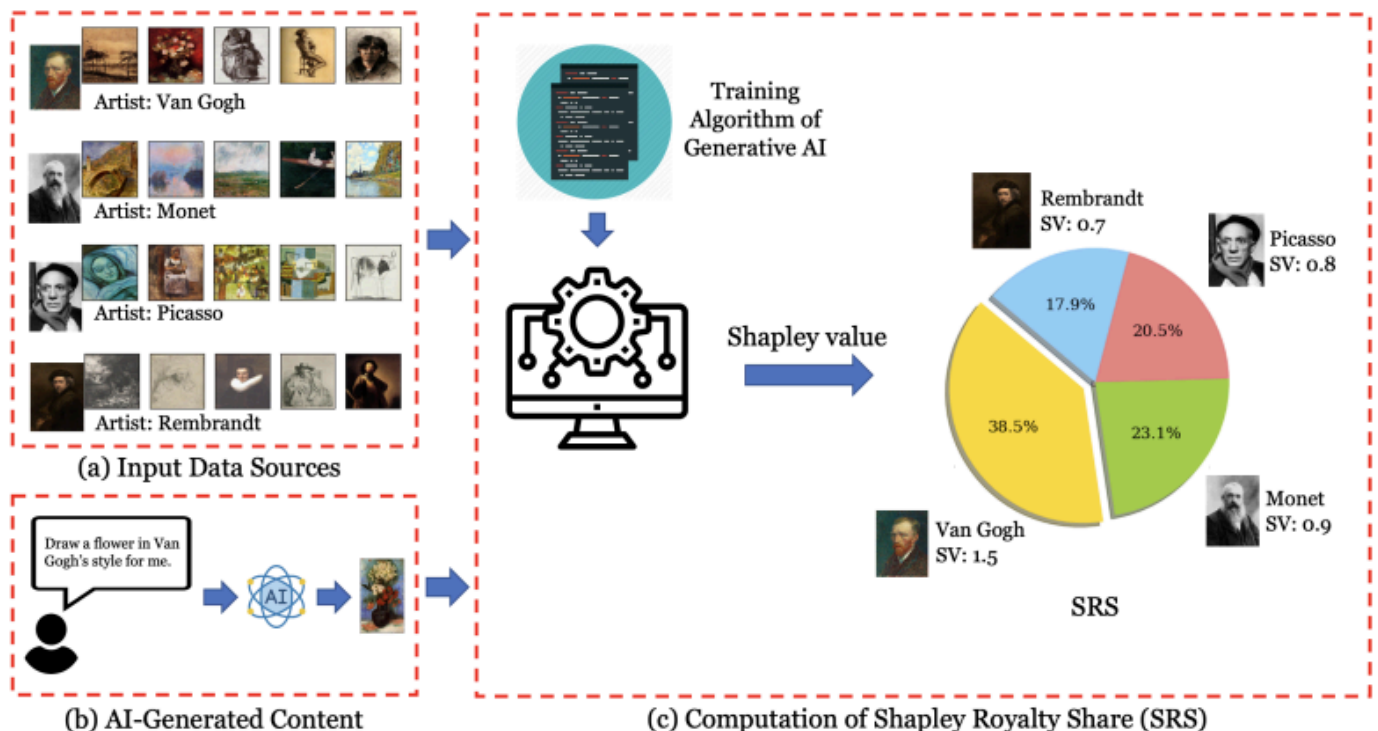
## 4 - Solution idéale

En théorie des jeux, Lloyd Shapley a introduit l'idée en 1953 qu'il était possible, dans un jeu coopératif, de calculer une répartition équitable des gains entre les joueurs. On considère par exemple un écosystème marin avec toutes ses espèces vivantes - algues, anémones, coquillages, étoiles de mer – et on mesure l'utilité de chaque espèce par sa contribution à la vie dans l'écosystème. Quand on retire les étoiles de mer, 50% des autres espèces disparaissent, contre 10% quand on retire les anémones. La valeur de Shapley des étoiles de mer est cinq fois supérieure à celle des anémones.

La solution idéale pour rémunérer les auteurs fait consensus chez les chercheurs. Il s'agit de calculer la valeur de Shapley pour chaque auteur et d'effectuer la répartition des rémunérations au pro rata de cette valeur<sup>29</sup>.

Dans le cas de l'IA générative, la répartition équitable des rémunérations des œuvres d'entraînement devrait idéalement se faire en fonction de l'utilité des œuvres d'entraînement pour les contenus AI-générés, par auteur. Dans la figure ci-dessous, la valeur de Shapley de Van Gogh est de 1,5 soit 38,5% de la rémunération totale.

Valeur de Shapley



En pratique, le coût de calcul de la valeur de Shapley est prohibitif, ce qui pose la question de la faisabilité technique d'une solution de rémunération équitable.

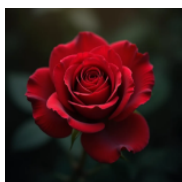
## 5 – Démonstrateur

### Objet

Le lecteur trouvera un simulateur en ligne à l'adresse  
<https://huggingface.co/spaces/TheFrenchDemos/genai-reward>








Nous pensons qu'il est aujourd'hui techniquement et économiquement possible de calculer un score d'attribution à des contenus d'entraînement pour répartir équitablement une rémunération entre les ayants-droits de ces contenus. Ce simulateur permettra à chacun de vérifier par lui-même ses propres hypothèses économiques.

Par exemple une image de rose a été générée avec le prompt « *A vibrant red rose in full bloom, macro photograph* » :



Le simulateur identifie en quelques secondes, parmi les dix millions de contenus de la base Open Images, les huit images originales qui ont le plus contribué à la formation du contenu généré. Il calcule un taux d'attribution, une rémunération, et fournit des informations d'identification du contenu original et de son auteur.

#### Attributions des images de roses

CONTENU	ATTRIBUTION	RÉCOMPENSE DU CONTENU (c€)	URL DU CONTENU	NOM DE L'AUTEUR	URL DE L'AUTEUR
	21%	0,63	<a href="https://www.flickr.com/photos/50093642@N03/4629227563">https://www.flickr.com/photos/50093642@N03/4629227563</a>	Yoko Nekonomania	<a href="https://www.flickr.com/photos/nekonomania/">https://www.flickr.com/photos/nekonomania/</a>
	18%	0,54	<a href="https://www.flickr.com/photos/22305657@N02/11733854773tt">https://www.flickr.com/photos/22305657@N02/11733854773tt</a>	Bruno Pinho	<a href="https://www.flickr.com/photos/brunopinho/">https://www.flickr.com/photos/brunopinho/</a>
	15%	0,45	<a href="https://www.flickr.com/photos/12692992@N00/170278141">https://www.flickr.com/photos/12692992@N00/170278141</a>	Dean + Barb	<a href="https://www.flickr.com/photos/barbanddean/">https://www.flickr.com/photos/barbanddean/</a>
	12%	0,36	<a href="https://www.flickr.com/photos/33140999@N07/3509399376">https://www.flickr.com/photos/33140999@N07/3509399376</a>	Michael Arrighi	<a href="https://www.flickr.com/photos/arrighi/">https://www.flickr.com/photos/arrighi/</a>
	10%	0,30	<a href="https://www.flickr.com/photos/55658968@N00/8730223835">https://www.flickr.com/photos/55658968@N00/8730223835</a>	Toshihiro Gamo	<a href="https://www.flickr.com/photos/dakiny/">https://www.flickr.com/photos/dakiny/</a>
	9%	0,27	<a href="https://www.flickr.com/photos/34226407@N04/8734050294">https://www.flickr.com/photos/34226407@N04/8734050294</a>	Edd Prince	<a href="https://www.flickr.com/photos/eddpics/">https://www.flickr.com/photos/eddpics/</a>
	9%	0,27	<a href="https://www.flickr.com/photos/41353201@N07/4267426012">https://www.flickr.com/photos/41353201@N07/4267426012</a>	Musicman Loxton	<a href="https://www.flickr.com/photos/dj-dwayne/">https://www.flickr.com/photos/dj-dwayne/</a>

	8%	0,24	<a href="https://www.flickr.com/photos/64574638@N04/6006728032">https://www.flickr.com/photos/64574638@N04/6006728032</a>	serbestruhpro	<a href="https://www.flickr.com/photos/serbestruhpro/">https://www.flickr.com/photos/serbestruhpro/</a>
---	----	------	---	---------------	---

En bas de page web, trois boutons permettent à l'utilisateur de télécharger :

- Ce rapport
- Le rapport des ayant-droits que l'exploitant pourrait générer et communiquer aux ayant-droits en application d'accords de règlement et rétrocession de la rémunération équitable
- Le rapport réglementaire prévu par l'AI act (art 53, 1, d), selon le formulaire proposé par le Higher Council for Literary and Artistic Property (CSPLA)

## Comment ça marche

Le simulateur dispose d'un bouton « advanced parameters » qui permet à l'utilisateur de régler les paramètres qu'il souhaite. Le tableau suivant recopie ceux qui sont fixés par défaut et indique les formules de calcul.

### Paramètres et valeurs d'une simulation

VARIABLE	PARAMÈTRE	COMPARABLE	CALCUL	FORMULE
Abonnement utilisateur	12 €/mois (1)	Adobe Firefly à partir de 4,99\$/mois Midjourney à partir de 10\$/mois Dall-e inclus avec ChatGPT Plus 20\$ Getty Edify 45€ pour 25 générations	S	
Contenus générés	60 /mois	Adobe Firefly plan de base 100/mois Midjourney plan de base 100/mois	G	
Valeur faciale par contenu généré			0,2 €	$V = S / G$
Part des titulaires de droits et les auteurs				$R\% = Rr\% + Ra\%$
- Titulaires des droits	10%	Musique (ASCAP, BMI, SACEM) 2 – 8% Image de stock 55 – 85%	Rr%	
- Auteurs	5%	Livres imprimés 5 – 15% Musique (interprète et compositeur) 15 – 30% Photographie de stock 15 – 45%	Ra%	
Rémunération des titulaires de droits et des auteurs par contenu généré			3 c€	$R = V \times R\%$

Attribution		21%	A% = Algorithme de similarité
		8	N = nombre de contenus attribués
Récompense du contenu de formation		0,63 c€	T = R x A%
Nombre d'auteurs	1.800		B
Utilisateurs	500.000	On estime entre 2 et 5 millions le nombre d'abonnés payants de Midjourney.	U
Récompense mensuelle pour l'ayant-droits et l'auteur		829 €	S x U x R% / B x A% x N
- Ayant-droits		276 €	S x U x Rr% / B x A% x N
- Auteur		553 €	S x U x Ra% / B x A% x N

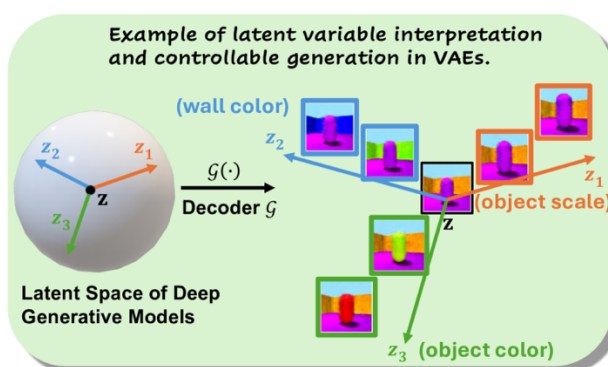
## Notion de similarité

Ce démonstrateur utilise un algorithme de similarité entre contenus d'entraînement et contenus générés, mettant en œuvre [DINOv2](#) et [FAISS](#). Intuitivement, deux images similaires se ressemblent.

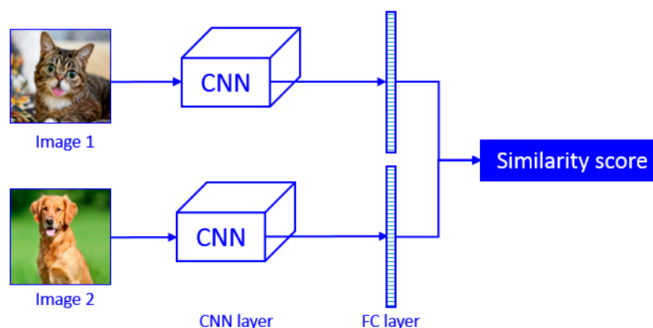
Pourtant, la « similarité » évoquée ici est une notion scientifique détachée de toute notion juridique de « contrefaçon », de « ressemblance » ou d'atteinte à une œuvre, définie en droit français comme « toute reproduction, représentation ou diffusion, par quelque moyen que ce soit, d'une œuvre de l'esprit en violation des droits de l'auteur, tels qu'ils sont définis et réglementés par la loi » (CPI, art. L. 335-3). Il n'y a pas nécessairement contrefaçon ou atteinte par la reproduction d'un élément d'une œuvre, une « contribution » en l'occurrence. De la même manière lorsque qu'un auteur crée, un élément peut entrer dans sa composition, sans qu'il n'y ait contrefaçon ou atteinte à un tiers. La rémunération qui découle de la « similarité » ne devrait donc pas paralyser l'action de l'auteur en cas de contrefaçon.

## Similarité vectorielle

Techniquement, chaque contenu généré ou original est analysé pour en extraire des informations mesurables : chat, chien, couleur et taille de l'avant-plan, type, style, taille, objets reconnus, mots-clés,... l'ensemble de ces centaines d'indicateurs d'utilités, environ 500 dans notre démonstrateur, forme un vecteur dans un « espace latent ».



Le degré d'attribution d'une œuvre (image 1) à un contenu généré (image 2) est calculée par un score de similarité entre les deux images, traitées par les deux réseaux neuronaux siamois. Plus les vecteurs résultants sont proches<sup>30</sup>, plus le score de similarité entre les images est élevé<sup>31</sup>.



## Nombre d'auteurs

La base Open Images a été formée à partir de celle de Flickr, en appliquant des filtres tels que la non-détection par Google Image Search, un nombre minimum de pixels, ou le dédoublement<sup>32</sup>. Open Images ne sélectionne aussi que les licences CC-BY, soit une image sur 30<sup>33</sup>. Comme Flickr propose par défaut la licence « tous droits réservés »<sup>34</sup>, le changement de licence n'est fait que par des auteurs avertis. Ceci suggère une forte concentration dans Open Images des contributeurs les plus actifs, ayant publié plus de 5.623 photos<sup>35</sup>, soit environ 1.800 auteurs. Cette hypothèse intuitive pourrait être vérifiée avec un comptage exact des auteurs de la base Open Images, que nous n'avons pas pu réaliser compte-tenu de la limite en vitesse de requête.

## Identification des auteurs

La base Open Images identifie les auteurs des images, et fournit la page de leur profil sur Flickr qui, en revanche, ne comprend pas d'adresse électronique ni postale complète.

Au-delà de notre démonstrateur et d'Open Images, les datasets publics donnent des éléments d'identification, de précision variable, mais qui peuvent être recoupés pour identifier chaque ayant-droits.

Les méthodes de recouplement sont surtout connues pour leur utilisation frauduleuse contre la vie privée. Dès 1997, la déclaration obligatoire de statistiques agrégées par un assureur du Massachussetts, avec des données électorales, avait permis d'identifier sans erreur possible le Gouverneur William Weld et de connaître son bilan de santé<sup>36</sup>.

Leur utilisation légale rend possible, dans la plupart des cas, l'identification complète des auteurs : par comparaison de graphes sociaux, par corrélation statistique grâce à une combinaison rare d'attributs (niche scientifique, heures d'activité, adresse IP), par reconstruction d'attributs individuels à partir de moyennes de groupe et d'attributs connus d'autres individus, etc<sup>37</sup>. Le tableau suivant indique des pistes d'identification des auteurs des contenus d'entraînement de GPT 4.



AI-generated by Adobe Firefly

#### Identification potentielle des auteurs de datasets utilisés par GPT-4

	DATASET	ELEMENTS D'IDENTIFICATION	RECOUPEMENT DIRECT	IDENTIFIANT
<b>Identification directe</b>	Open Images	-	-	Nom, prénom, fiche de profil
	Project Gutenberg			Nom, prénom, dates_naissance_et_mort
	arXiv	-	-	Nom, Initiale_prénom
	PubMed38	-	-	Id_author
<b>Identification indirecte</b>	CommonCrawl	Domaine	Whois	Registrant_contact (name, organization, street, city, state, phone)
	BookCorpus	Plain_text <sup>39</sup>	Smashwords	Nom, prénom
	WebText	Plain_text + search = « written by » <sup>40</sup>	API Amazon	Nom, prénom
	StackOverflow	HTML <sup>41</sup>	Stackoverflow.com	User_ID <sup>42</sup> , network_profile <sup>43</sup>
<b>En attente de revendication</b>	Wikipedia	Pseudo_utilisateur	-	-
	Reddit	Pseudo_utilisateur	-	-
<b>Autre</b>	CommonVoice44	(stats) <sup>45</sup>	-	-
	Google Books Ngrams	(non applicable)	(na)	(na)
	GitHub	(non accessible)		



## 6 – Solutions existantes

### Synthèse

Le tableau suivant identifie quelques méthodes commercialisées ou expérimentées, au regard de notre démonstrateur :

#### Méthodes de valorisation des contenus d'auteurs

VALORISATION	MÉTHODE DE VALORISATION	DÉMONSTRATEUR	SOCIÉTÉS COMMERCIALES	CHERCHEURS
Contenu généré	Coût d'abonnement / quantité	Controv3rse		
Dataset	Similarité d'ensembles de contenus			Adobe
	Shapley			Open AI
	In-run data Shapley			Open AI
	Fréquence d'accès au site web, type de données		Tollbit <sup>46</sup>	
Œuvre d'entraînement	Similarité de contenus	Controv3rse	Bria (images), Prorata Gist.ai <sup>47</sup> (textes)	Open AI
	Modélisation de l'apprentissage			
	Personnalisation du modèle			Adobe
	Désapprentissage			Adobe
	Tatouage			Adobe
	Shapley fédéré			Huawei
Auteur	Pay to train		Shutterstock, Getty, Audiosparx, Canva	
	Influence de l'ensemble des oeuvres		Sureel.ai <sup>48</sup> (musique)	
	Identification des ayant-droits	Controv3rse		

## Solutions commercialisées

### Entreprises rémunérant les auteurs de contenus d'entraînement

SOCIÉTÉ	CAPITAL-RISQUE	CHIFFRE D'AFFAIRES	MÉTHODE	MEDIA	FOURNISSEUR DE CONTENUS
Adobe (Firefly <sup>49</sup> )	-	\$ 21 billion	Pay-to-train	Image	
Canva	-	\$ 2 billion	Pay-to-train	Image	
Getty	-	\$ 917 million	Pay-to-train	Image	
Shutterstock	-	\$ 875 million	Pay-to-train	Image	
ProRata (Gist.ai <sup>50</sup> )	\$ 25 million	-	Similarité de contenu	Texte	Sky News, Time, Financial Times, The Atlantic, Axel Springer, The Guardian, dmg media, Universal Music, Fortune <sup>51</sup>
Bria	\$ 24 million <sup>52</sup>	-	Similarité de contenu	Image	Shutterstock, Getty Images, Alamy, Superstock
Tollbit	\$ 24 million <sup>53</sup>	-	Fréquence d'accès au site web	Texte	Penske Media Corporation, Time, Mumsnet, Trusted Media Brands, Candr Media Group, AdWeek
Sureel.ai	-	-	Influence de l'ensemble des oeuvres	Image, musique	
Audiosparx	-	-	Pay-to-train	Musique	

### Méthode Pay-to-train

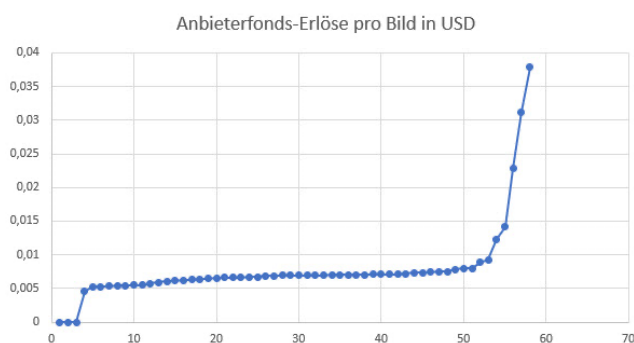


AI-generated by DALL-e

Les banques de contenus privées d'Adobe, Canva, Getty et Shutterstock, répartissent le prix des datasets qu'elles commercialisent en application de la méthode « pay to train ». Les auteurs sont rémunérés en fonction de leur volume de contribution à ces datasets.

La valeur « pay-to-train » est calculable à partir d'indicateurs d'utilités divers. Shutterstock rémunère par exemple chaque photo à 0,0050 \$ par an, auquel s'ajoute jusqu'à 0,0038 \$ par photo et par an en fonction de la quantité de métadonnées et autres indicateurs<sup>54</sup>.

Valeur Pay-to-train : rémunération des auteurs par Shutterstock. Echantillon N = 58 (en \$/image)<sup>55</sup>



L'avantage de ce modèle est sa simplicité de calcul. Il est opérationnel depuis deux ans<sup>56</sup>, et semble bien accepté par les ayant-droits.

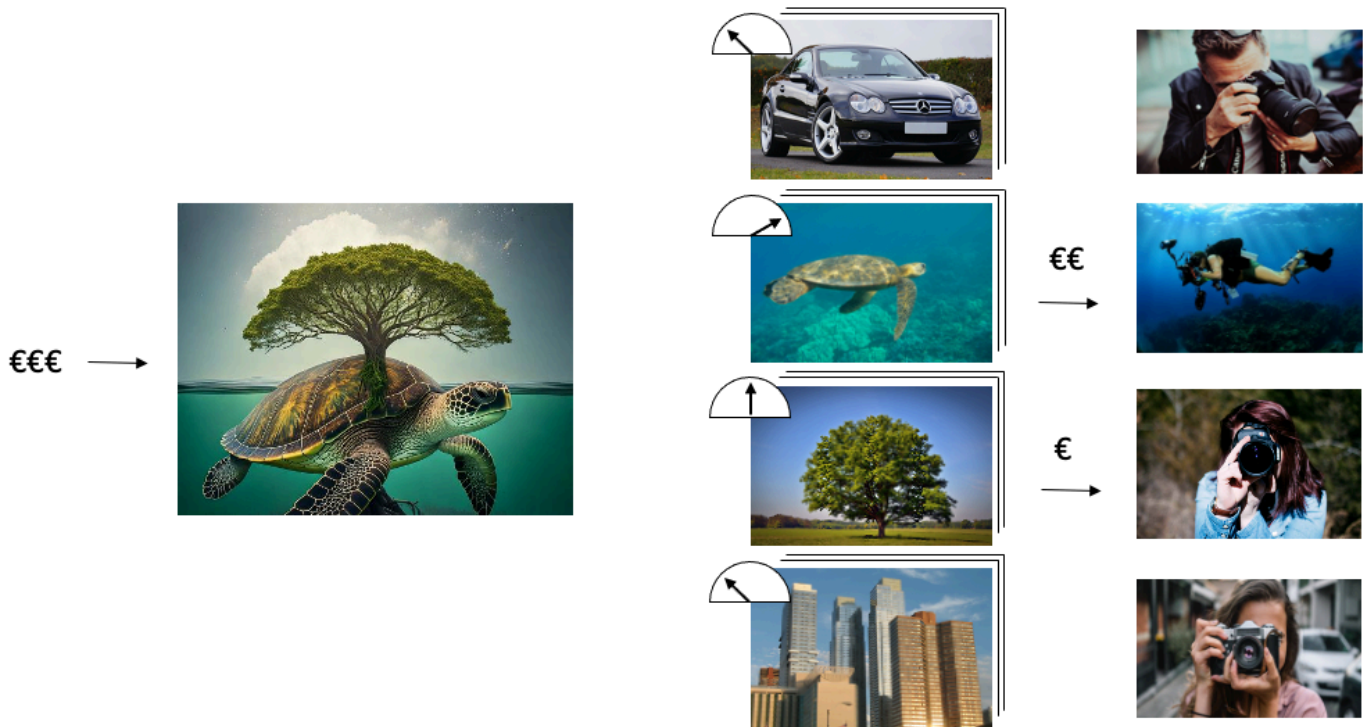
En revanche le niveau d'équité entre auteurs présente deux biais :

- (1) la qualité du contenu n'est pas prise en compte : un auteur ayant contribué avec 200 photos reçoit, toutes choses égales par ailleurs, une rémunération deux fois supérieure à un contributeur ayant contribué avec 100 photos
- (2) l'utilité pour le contenu généré n'est pas prise en compte : Van Gogh ne reçoit pas davantage que Picasso si les utilisateurs insèrent plus souvent « à la manière de Van Gogh » dans leurs prompts.

## Méthode hybride Similarité + Pay-to-train

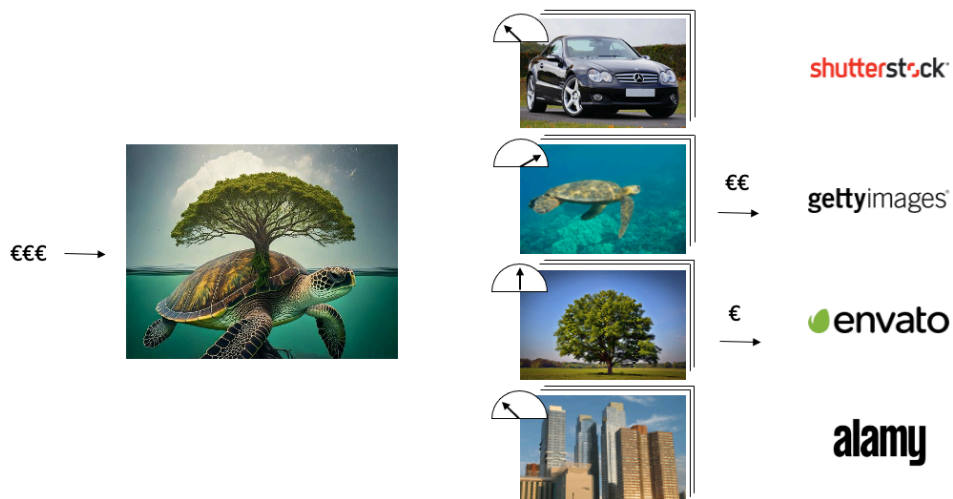
La startup israélo-américaine Bria, qui collabore avec les grandes banques d'images privées (Shutterstock, Getty), est intégrée par les Big Tech (Microsoft, Amazon, Nvidia). En première approche, Bria utilise comme notre démonstrateur une méthode de similarité, afin de rémunérer les auteurs de manière proportionnelle à la ressemblance ainsi mesurée.

Mesure des similarités entre contenu généré et photos originales pour rémunérer les auteurs



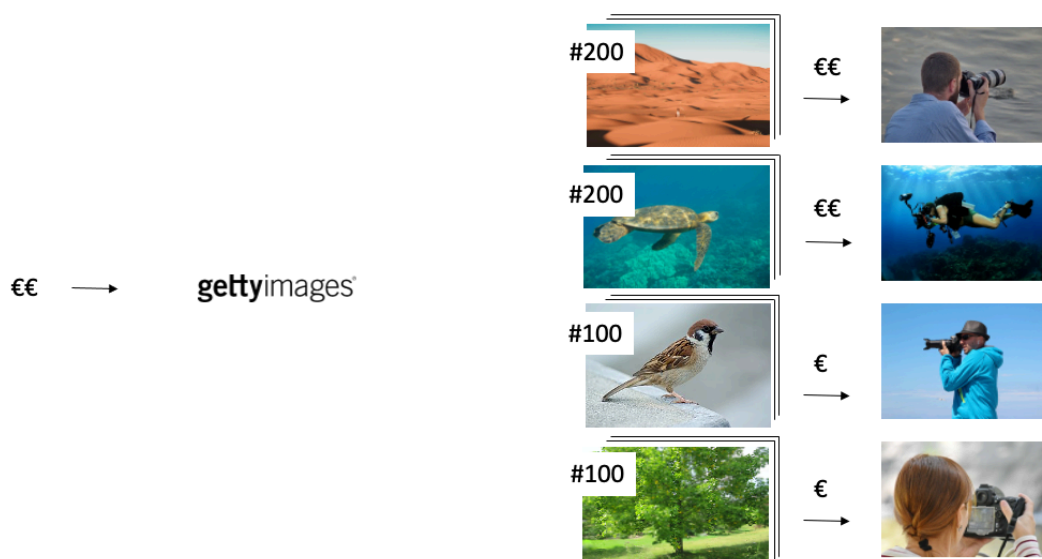
Plus précisément, il semble que la somme des scores de similarité soit calculée par banque d'images (Shutterstock, Getty, Alamy...), ce qui provoque une première rémunération des banques proportionnelle à ces sommes.

Mesure des similarités entre contenu généré et photos originales pour rémunérer les banques d'images<sup>57</sup>



Ces banques répartissent ensuite cette première rémunération entre les auteurs, en fonction du nombre individuel de leurs œuvres contribuant aux datasets vendus, selon la méthode « pay to train ».

Nombre de photos pour rémunérer les auteurs



D'autres affinages sont apportés, les images assorties de métadonnées de qualité reçoivent ainsi un bonus.

La méthode de Bria se fait donc en deux temps :

- D'abord grâce à un algorithme de similarité entre images d'entraînement et images générées
- Ensuite grâce à un algorithme de contribution quantitative des auteurs aux datasets vendus (« pay-to-train »)

Cette solution hybride est :

- économiquement compatible avec le modèle d'affaires des IA génératives
- acceptée depuis deux ans par les banques d'images et les auteurs
- en cours d'extension à d'autres types de contenus : 3D, musique, voix, image animée, vidéo<sup>58</sup>

Il est possible que la startup ProRata, en particulier, ait adopté cette méthode hybride et la propose aux media qui ont déjà des accords avec des exploitants d'IA<sup>59</sup>. Il s'agit de « prendre le résultat de l'IA générative, qu'il s'agisse de texte, d'image, de musique ou d'un film, et le décomposer en composants, pour déterminer d'où ils viennent, puis attribuer un pourcentage à chaque titulaire de droits d'auteur, et les payer en conséquence »<sup>60</sup>. Les brevets publics de sa société-mère<sup>61</sup> ne donnent pas davantage de précision.

Cette méthode hybride présente encore des défauts :

- il n'est pas prouvé que cette solution s'adaptera à tous les types de contenus, par exemple les jeux
- rien ne garantit que la similarité n'est pas fortuite. En d'autres mots, un auteur pourrait être rémunéré alors que ses œuvres n'ont pas été utiles pour les contenus générés, et réciproquement
- le degré de contribution d'une image originale est mesuré de manière approximative. Dans l'exemple ci-dessus, le photographe du désert est rémunéré autant que celui de tortues marines, car leur nombre de contributions au dataset de Getty Images est identique, alors que l'utilité des photos de désert au contenu généré est nul

On notera que :

- même si la répartition de la rémunération pour un contenu généré particulier est approximatif, son équité s'améliore probablement avec la quantité de contenus générés.

- l'identification individuelle des auteurs par Bria n'est pas nécessaire, mais seulement l'identification des datasets d'entraînement
- l'algorithme de répartition lui-même n'est pas public. Il est possible que les auteurs réclament dans l'avenir davantage de transparence, mais il est probable aussi, à l'inverse, qu'une dose de secret soit nécessaire pour éviter des biais opportunistes.
- on ne connaît pas non plus le taux moyen de rémunération des auteurs par rapport au chiffre d'affaires des contenus générés qui, à notre avis, pourrait quant à lui être publié.

## Méthode de la similarité par dataset

La startup Sureel fournit aux musiciens un outil qui mesure le niveau d'influence de leur travail, mesuré en pourcents, sur un morceau généré par l'IA.

L'algorithme, qu'il décrit dans son brevet '949<sup>62</sup>, prévoit la rémunération d'un auteur en fonction de l'influence de l'ensemble de ses œuvres, par exemple mesurée par la différence entre les poids du modèle de base et ceux de son réglage fin par cet ensemble. Cette variante vise à diluer le biais de la similarité grâce à un plus grand nombre d'œuvres comparées.

## Méthodes en cours d'expérimentation

Au-delà de ces quelques variantes, les chercheurs rivalisent d'ingéniosité pour tester des méthodes nouvelles :

### Recherche de modèles de rémunération équitables

ENTREPRISE	MÉTHODE	ÉQUIPE DE RECHERCHE	DATE	MOTIVATION
OPEN AI Princeton University Columbia University Harvard University University of Pennsylvania University of California Berkeley	Shapley	<b>An Economic Solution to Copyright Challenges of Generative AI</b> <sup>63</sup> Jiachen T. Wang, Zhun Deng, Hiroaki Chiba-Okabe, Boaz Barak, Weijie J. Su	14 avril 2024 9 septembre 2024 (v4)	Les valeurs de Shapley offrent une approche fondée sur des principes de répartition des gains en fonction de l'utilité.  L'utilitaire permet de mesurer dans quelle mesure les sources de données sont responsables de la génération du contenu. La valeur est faible si le modèle contrefactuel est peu susceptible de générer le même contenu que le modèle déployé, et vice versa.  L'utilité peut être interprétée comme la compensation totale que tous les membres méritent collectivement pour fournir leurs données afin de former le modèle d'IA générative.

ENTREPRISE	MÉTHODE	ÉQUIPE DE RECHERCHE	DATE	MOTIVATION
Virginia Tech Massachusetts Institute of Technology	In-run data Shapley	<b>Data Shapley in One Training Run</b> <sup>64</sup> Jiachen T. Wang, Prateek Mittal, Dawn Song, and Ruoxi Jia	29 juin 2024	Boaz Barak, professeur à Harvard en charge de la sécurité (« superalignement » <sup>65</sup> ), également chez OpenAI, ne perd pas de vue la méthode idéale de Shapley, même si cela signifie faire des concessions sur la précision.  Son équipe de chercheurs montre en grandeur nature comment calculer la valeur de Shapley d'ensembles de données comme Wikipédia ou Github pour un modèle d'IA à usage général formé à partir d'un ensemble <sup>66</sup> comparable à celui de GPT.  Selon cette équipe, aligner la rémunération sur les contributions quantifiables de chaque jeu de données garantit la meilleure équité dans la répartition des royalties <sup>67</sup> . Elle indique sa préférence pour la variante In-run data Shapley, moins coûteuse.
	Training modeling	<b>TRAK: Attributing Model Behavior at Scale</b> <sup>68</sup> Sung Min Park, Kristian Georgiev, Andrew Ilyas, Guillaume Leclerc, Aleksander Madry MIT	3 avril 2023	Aleksander Madry, professeur au MIT, est responsable du raisonnement IA chez OpenAI. La solution Trak s'appuie sur une représentation approximative du modèle d'IA pour décrire son évolution au cours de l'apprentissage et relier les prédictions aux données d'entraînement qui les ont influencées.  Cette solution s'est avérée utile avec les classificateurs d'images formés sur ImageNet, les modèles vision-langage (CLIP) et les modèles linguistiques (BERT et mT5).
ADOBE Carnegie Mellon University UC Berkeley	Dataset similarities	<b>Evaluating Data Attribution for Text-to-Image Models</b> <sup>69</sup> Sheng-Yu Wang, Alexei Efros, Jun-Yan Zhu, Richard Zhan	8 août 2023	Comme la méthode de similarité risque de révéler des ressemblances fortuites, ce biais peut être dilué en étendant la méthode à deux ensembles d'images, d'entraînement et générées. Les chercheurs d'Adobe calculent ainsi la contribution du premier ensemble, le dataset, à un ensemble de contenus synthétiques
	Watermarking	<b>ProMark: Proactive Diffusion Watermarking for Causal Attribution</b> <sup>70</sup> Vishal Asnani, John Colomosse et al.	14 mars 2024	Le tatouage démontre la causalité entre un contenu d'entraînement et un contenu synthétique. Cela permet de fonder avec confiance l'attribution légale ou les paiements aux créateurs
	Unlearning	<b>Data Attribution for Text-to-Image Models by Unlearning Synthesized Images</b> <sup>71</sup> Sheng-Yu Wang, Aaron Hertzmann, Alexei A. Efros, Jun-Yan Zhu, Richard Zhang	juin 2024	Cette méthode désapprend l'image synthétisée, en entraînant les images oubliées par proxy, en identifiant celles qui présentent des écarts de perte significatifs après le processus de désapprentissage, et en les étiquetant comme influentes.



ENTREPRISE	MÉTHODE	ÉQUIPE DE RECHERCHE	DATE	MOTIVATION
HUAWEI University of British Columbia Simon Fraser University KTH Royal Institute of Technology	Federated Shapley	<b>Improving Fairness for Data Valuation in Horizontal Federated Learning</b> <sup>72</sup> Zhenan Fan, Huang Fang, Zirui Zhou, Jian Pei, Michael P. Friedlander, Changxin Liu, Yong Zhang	23 mai 2022	La motivation des propriétaires de données dépend en partie de l'équité de la collaboration et de la récompense dans l'apprentissage fédéré.

## Variantes et évolutions

Dans toutes les méthodes évoquées :

- les paramètres de rémunération sont réglables en fonction d'arbitrages sur les indicateurs d'utilité : taille, date, similarité, nombre de métadonnées, succès commercial indépendant, etc.

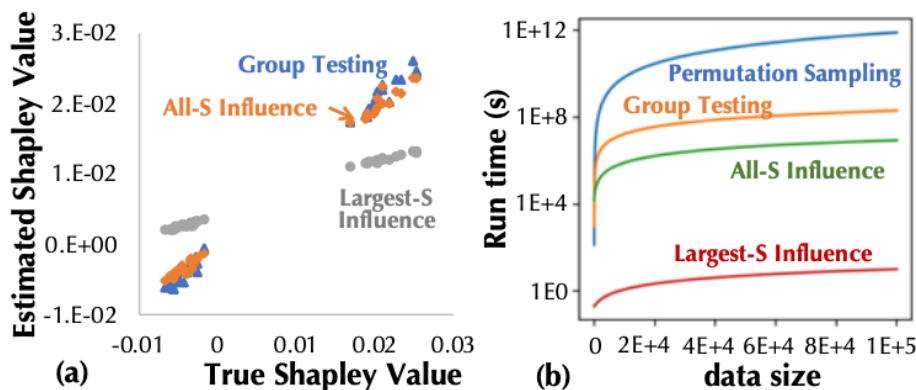
Indicateurs d'utilité

CHERCHEUR, ENTREPRISE	MÉDIAS	INDICATEURS D'UTILITÉ
Piao, Chen et al. <sup>73</sup>	Image	Vecteurs
Wang, Deng et al. <sup>74</sup>	Encyclopédie, Publication académique, Image, Code	Modèle contrefactuel (S) Contenus générés par le modèle x(gen) et le modèle contrefactuel
Wang, Mittal et al. <sup>75</sup>	Encyclopédie, Publication académique, Image, Code, Email, Forum	Ensemble de points de données de validation de maintien Z( val )
Deng, Zhang et al. <sup>76</sup>	Musique	Vitesse moyenne des événements Hauteur moyenne de tous les événements Somme des écarts de temps de tous les événements
Shutterstock + Bria	Image, image vectorielle	Score de similarité Chiffre d'affaires par jeu de données Volume de données par jeu de données et par auteur Volume des métadonnées Rôle joué dans le développement des modèles originaux Rôle joué par le biais de paiements de redevances liés à une future activité de licence générative
Adobe Stock + Bria	Image, image vectorielle	Score de similarité Admissibilité des images Date de l'image Nombre de téléchargements



- l’algorithme est réglable entre précision et vitesse d’exécution. La valeur de Shapley, en particulier, peut être calculée
  - avec un nombre arbitraire de contenus générés
  - avec une granularité arbitraire (par dataset, sous-dataset, jusqu’à la donnée individuelle)
  - avec des fonctions approchantes<sup>77</sup>

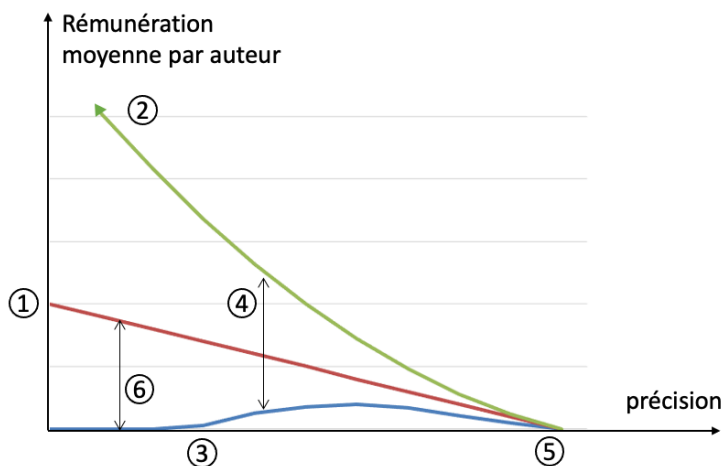
Trois fonctions approchantes de la valeur de Shapley<sup>78</sup>



A ce stade de l’état de l’art, le champ d’investigation est suffisamment large pour espérer trouver des solutions de rémunération, par usage, qui optimisent le rapport entre équité, coût et précision.

Impact du coût de la précision sur la rémunération.

- (1) A précision nulle, la rémunération par auteur est égale à la collecte totale divisée par le nombre d’auteurs.
- (2) A précision nulle, la rémunération potentielle maximale d’un auteur est égale à la collecte totale
- (3) A précision nulle, la rémunération potentielle minimale d’un auteur est égale à zéro
- (4) Ecart d’incertitude
- (5) Quand le coût de la précision dépasse la collecte totale, la rémunération et l’écart d’incertitude sont nuls
- (6) A collecte constante, la rémunération moyenne par auteur diminue avec la précision



## 7 – Acceptabilité

**JDN**  
JOURNAL DU NET

**#FairlyShare contre le péché originel d’Internet** – septembre 2014, par Vincent Lorphelin, Gilles Babinet et al.

Ce qu’ils dénoncent ? Les plate-formes qui reposent sur l’industrialisation du travail gratuit, en dehors des règles de l’appel d’offre, du concours ou du bénévolat. Ils emboîtent ainsi le pas à [...] la levée de boucliers qui a suivi le rachat du journal de crowdsourcing Huffington Post par AOL pour 315 millions de dollars. [Lire](#)

La méthode « pay-to-train » est actuellement la plus répandue pour rémunérer les auteurs dont les œuvres sont utilisées pour l’entraînement des modèles d’IA à usage général. Malgré une transparence et une équité partielles, elle n’a pas suscité pour l’instant de procès ni de pétitions, comme on en avait vu par exemple lors de la prise de conscience du « travail gratuit » ou de la cession du Huffington Post (voir encadré).

Banques de media utilisant le « pay-to-train » et modèles d’IA utilisant leurs datasets

DATE	MEDIA	FOURNISSEUR DE DATASET	FOURNISSEUR DE MODELE D’IA
10/25/22	Image, photo	Shutterstock	OpenAI
01/12/23	Image, photo	Shutterstock	Meta, Google, Amazon, Apple
03/12/23	Image, photo	Shutterstock	LG (Exaone)
03/21/23	Image, photo	Getty Images	Nvidia
05/29/23	Avatar	Ascendant Art	
07/11/23	Image, photo	Shutterstock	OpenAI
08/08/23	Image, photo	Shutterstock	Nvidia (Picasso)
09/07/23	Image, photo	Getty Images Alamy Bria	

DATE	MEDIA	FOURNISSEUR DE DATASET	FOURNISSEUR DE MODELE D’IA
09/13/23	Music	AudioSparx	Stability AI
09/13/23	Image, photo	Adobe Stock (Firefly)	
09/26/23	Image, photo	Getty Images (Getty Gen AI)	
10/04/23	Image, photo	Canva (Magic Studio)	Open AI (Dall-e) Google (Imagen)
02/21/24	Image, photo	Superstock <sup>79</sup> Envato	
04/11/24	Image, photo	EyeEm Freepick Photobucket	
06/27/24	Music, Voice	Universal Music* Sony Entertainment Warner Records	Google (Youtube)
07/30/24	3D	Shutterstock <sup>80</sup>	Nvidia (Edify)

\*expérimentation

Les contrats d'achats de datasets auprès des banques de media ont été accompagnés de discours qui valorisent la rémunération « respectueuse, responsable, sécurisée et équitable » des auteurs.

### Discours sur l'équité du pay-to-train

PLAT-FORME	STOCK ET DATASET	DECLARATIONS
Nvidia	Getty, Shutterstock, Adobe	Nvidia : « Nous avons créé une plateforme qui permet à nos partenaires de s'entraîner à partir de données sous licence, par exemple auprès de Getty, Shutterstock ou Adobe », a déclaré Huang. « Ils respectent les propriétaires de contenu. Les données d'entraînement proviennent de cette source et, quels que soient les avantages économiques qui en découlent, ils peuvent revenir aux créateurs. » <sup>81</sup>
	Adobe	Adobe : nous développons une IA générative de manière responsable, avec les créateurs au centre <sup>82</sup> . Adobe : créez en toute confiance en sachant que votre résultat est de la plus haute qualité, construit avec des modèles conçus pour être sûrs pour une utilisation commerciale. <sup>83</sup> .
	Getty	Getty : Generative AI par iStock est un outil d'IA générative abordable et commercialement sûr [...] sans crainte que quelque chose de légalement protégé se soit glissé dans l'ensemble de données et puisse se retrouver dans leur travail <sup>84</sup>
Meta	Shutterstock	Shutterstock : une expertise dans la création d'un écosystème à grande échelle qui rémunère et connecte les contributeurs aux créateurs. <sup>85</sup> .
LG Exaone	Shutterstock	Shutterstock : La propriété du contenu généré par l'IA ne peut pas être attribuée à un individu et doit plutôt rémunérer les nombreux artistes qui ont participé à la création de chaque nouveau contenu. Nos clients peuvent acquérir des licences d'images IA en toute sécurité et légalement pour leurs propres besoins, sans souci. Nous rémunérons également les contributeurs dont les œuvres entraînent nos modèles, afin que vous puissiez également utiliser notre IA en toute bonne conscience. Notre générateur d'images IA reste la fonctionnalité sûre, facile à utiliser et conçue de manière responsable que vous connaissez et aimez <sup>86</sup>
		LG : Le modèle d'IA est formé avec des millions d'images haute résolution et de métadonnées de Shutterstock et convertira les invites textuelles en images. Les contributeurs dont les œuvres ont été utilisées pour former le modèle seront rémunérés via le Fonds des contributeurs de Shutterstock et seront également rémunérés chaque fois qu'un nouveau contenu génératif utilisant leur propriété intellectuelle sera créé et concédé sous licence par les clients <sup>87</sup> .
Open AI	Shutterstock	OpenAI : Les données que nous avons obtenues sous licence de Shutterstock ont été essentielles à la formation de DALL-E. Veiller à ce que l'économie des créateurs continue d'être dynamique est une priorité importante pour OpenAI . Les écrivains, les artistes, les compositeurs et les autres créateurs ont contribué de manière incommensurable aux sociétés tout au long de l'histoire de la civilisation [...] OpenAI ne veut pas remplacer les créateurs <sup>88</sup> .
	Canva	Canva : le convertisseur de texte en image de Magic Media, DALL·E d' OpenAI et Imagen de Google Cloud [...] garantissent que leurs modèles d'IA sont équitables et éthiques. « Nous partageons la valeur que nous développons dans Canva avec les créateurs qui nous accompagnent depuis toutes ces années. Avec le nouveau Creator Fund, ils ont accès à un pool de redevances qui correspond aux créations d'IA. » <sup>89</sup>
Google Cloud	Canva Shutterstock	Google Cloud : Shutterstock est également devenu un innovateur de premier plan en étant le premier à lancer un générateur d'images IA d'origine éthique, désormais amélioré avec Imagen sur Vertex AI.
Google Youtube	UMG	YouTube, filiale de Google, tente de conclure des accords avec les majors de la musique pour l'utilisation de l'IA. Il veut s'assurer que « les détenteurs de droits musicaux soient rémunérés pour leurs contributions aux données d'entraînement. UMG et YouTube sont en pourparlers pour obtenir des licences sur les voix et les mélodies des artistes pour entraîner les modèles d'IA ». <sup>90</sup>
Amazon SageMaker	Alamy	Alamy : Au cœur de l'éthique du programme Fair Diffusion se trouve l'engagement de rémunérer équitablement les photographes, les artistes et les créateurs <sup>91</sup> .
	Getty	Getty : nous adhérons aux normes éthiques les plus élevées et respectons les droits de propriété intellectuelle et de confidentialité des autres <sup>92</sup> .
	Envato	

PLAT-FORME	STOCK ET DATASET	DECLARATIONS
Sony		<p>Le président de Sony a déclaré : « Les nouveaux produits et entreprises créés à partir de la musique doivent être développés avec le consentement du propriétaire et une rémunération et un crédit appropriés. Il est essentiel de comprendre pourquoi la formation des modèles d'IA est effectuée, quels produits seront développés en conséquence et quel est le modèle commercial qui permettra de monétiser l'utilisation du travail de l'artiste »<sup>93</sup>.</p> <p>« Si un modèle d'IA générative est formé sur la musique dans le but de créer de nouvelles œuvres musicales qui rivalisent sur le marché de la musique, alors la formation n'est pas une utilisation équitable. Dans ce cas, la formation ne peut se faire sans le consentement, le crédit et la compensation des artistes et des titulaires de droits. »<sup>94</sup></p>

## 8 – Brevets

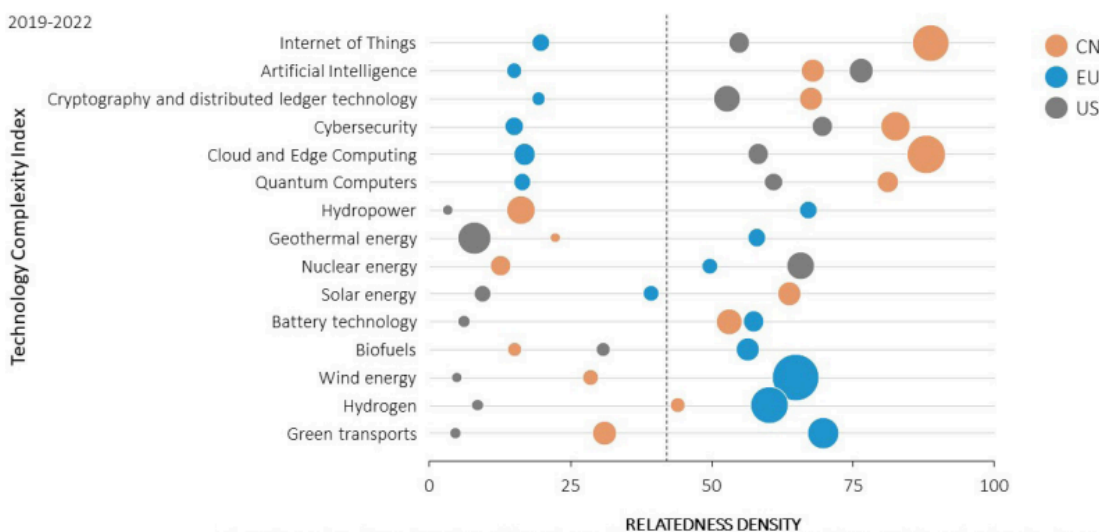
Lors de la rédaction de ce rapport, nous avons identifié les brevets suivants :

Brevets relatifs à la rémunération des oeuvres d'entraînement de l'IA générative

TITRE	REF	PAYS DU DÉPOSANT	DÉLIVRÉ	PIONNIER
Systems and methods of creative work collaborative systems	US9100444	U.S	✓	✓
Stakeholder equity valuation in collaborative projects	US10776889	U.S	✓	✓
Attributing generated visual content to training examples	US12033372	U.S	✓	
Model-based attribution for content generated by an artificial intelligence	US12013891	U.S	✓	
Generating author vectors	US10599770	U.S	✓	
Input-based attribution for content generated by an artificial intelligence	US2024419949	U.S		
Output-based attribution for content generated by an artificial intelligence	US2024419720	U.S		
Fee Distribution System and Method	US2023401550	U.S		
Generation, usage and attribution of content	WO2024166091	U.S		
Facilitating generation and usage of visual content	WO2023084512	U.S		
Federal learning-based online client selection method and system, medium and equipment	CN118446287	China	✓	
Image index generation based on similarities of image features	CN103207879	China	✓	
Method and system for evaluating contribution degree of image local features to overall features	CN111652239	China	✓	
Method and system for quantifying client contributions in federated learning	WO2023036184	China		
Federal learning mobile device distribution data processing method based on Shapley value	CN114912626	China		
Federal learning excitation method based on cooperation Sharpley value	CN115965092	China		
Methods and systems for federated learning with local predictors	WO2024031564	China		
Method and collaboration system	US 9723059	E.U	✓	✓
Method for an explainable autoencoder and an explainable generative adversarial network	US11948083	E.U	✓	
Architecture for explainable reinforcement learning	US11455576	E.U	✓	
Multistage contribution traceability in collective creation environment	US2015363743	E.U		
Method and device for rewarding collaborative creation	EP2958068	E.U		

Cette liste montre que, pour la rémunération des ayants-droits, les européens ne sont pas distancés par la Chine et les Etats-Unis, ce qui contraste avec leur position d'outsider<sup>95</sup> pour l'intelligence artificielle en général et l'IA générative en particulier<sup>96</sup>.

### The EU's position in complex (digital and green) technologies



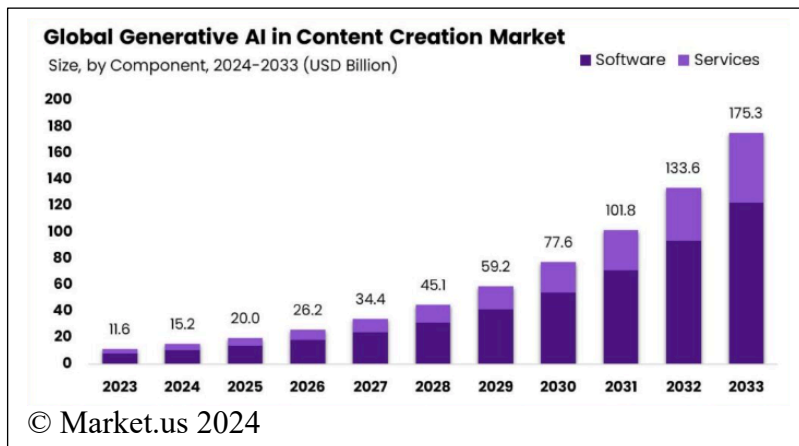
**Table 7 Top patent owners in GenAI modes (companies), 2014–2023**

	Images/ video	Text	Speech/ voice/ music	3D image models	Molecules/ genes/ proteins	Software/code	Other modes
Tencent Holdings (China)	607	565	551	102	57	41	464
Ping An Insurance (China)	262	600	599	26	33	24	223
Baidu (China)	395	465	441	81	26	18	166
IBM (US)	101	274	168	13	37	42	132
Alibaba Group (China)	142	213	144	36	2	10	133
Samsung Electr. (Republic of Korea)	173	140	226	34	10	5	57
Alphabet/Google (US)	138	107	200	25	15	15	72
ByteDance (China)	173	82	112	21	6	7	90
Microsoft (US)	78	194	151	18	6	22	42
BBK Electronics (China)	125	58	117	12	0	4	116
Netease (China)	78	76	80	37	0	11	102
NTT (Japan)	37	100	169	11	0	3	62
Huawei (China)	104	84	72	24	1	4	102
China Mobile (China)	79	67	83	10	0	7	89
State Grid (China)	75	54	31	6	1	2	144
Adobe (US)	190	79	34	30	0	9	20
Sony Group (Japan)	87	35	105	42	0	1	42
Siemens (Germany)	124	25	8	10	3	4	59
Ant Group (China)	45	82	29	3	0	6	59
Industrial and Commercial Bank of China (China)	40	62	49	5	0	16	48

Source: WIPO, based on patent data from EconSight/IFI Claims, April 2024.

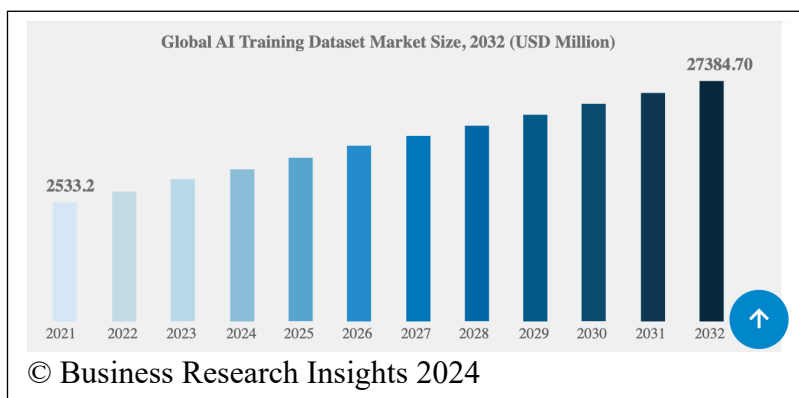
# Perspectives

## 9 – Marché



Pour le seul segment de la créativité, la taille du marché mondial de l’IA générative est estimée à 11,6 milliards USD en 2023 et dépassera 100 milliards USD d’ici 2031<sup>97</sup>.

Pour disposer d’ordres de grandeur, un taux de rémunération équitable de 15% représenterait davantage que la collecte actuelle des OGC (Organismes de Gestion Collective)<sup>98</sup> pour les droits artistiques (musique, audiovisuel, spectacle vivant, littérature et arts visuels).



Les entreprises spécialisées dans les datasets de formation à l’IA organisent la collecte, l’annotation, la conservation des données et leur adéquation aux besoins des algorithmes d’apprentissage.

Le marché mondial des datasets de formation en IA était de 2,5 milliards de dollars en 2021 et devrait atteindre 27 milliards de dollars d’ici 2032<sup>99</sup>.

### Modèles de rémunération des auteurs de données de l’IA générative

MODÈLE ÉCONOMIQUE	SOCIÉTÉS	RÉMUNÉRATION DES AUTEURS
Scraping (1)	OpenAI, Midjourney, Stability, MistralAI	Non
Uber des données (2)	Mturk, ClickWorker, Gigwalk, Premise	Paiement sur commande
Banques de contenus (3)	Shutterstock, Getty, Canva, Adobe	Pay-to-train
Groupes de media (4)	Axel Springer, Financial Times, Stack Overflow, Reddit, Le Monde	Non
Sociétés d’auteurs (5)	the Authors’ Guild, Created by Human, Human Native AI, Fairly Trained, Dataset Providers Alliance	Rémunération équitable

(1) Scraping : DALL·E, par exemple, a d’abord été formé avec des centaines de millions d’images sous-titrées provenant d’Internet<sup>100</sup>. Cette pratique, majoritaire<sup>101</sup>, est à l’origine de nombreux procès<sup>102</sup>.

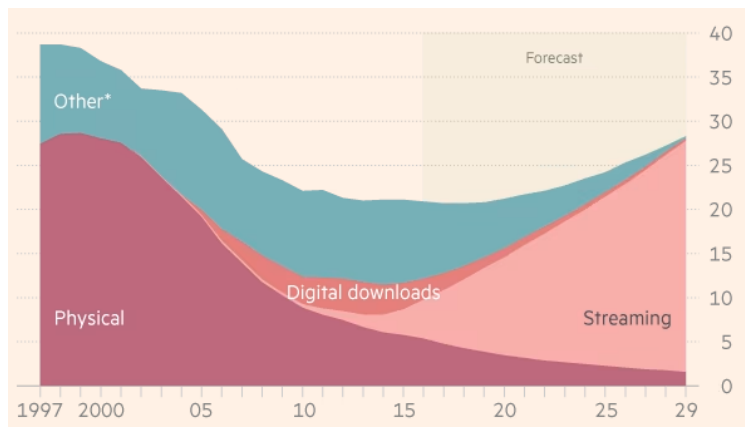


- (2) **Uber des données** : les créateurs de modèles d'IA ont commencé à couvrir les risques et à sécuriser les chaînes d'approvisionnement en données, à la fois via une industrie florissante de courtiers en données et par des accords avec les propriétaires de contenu qui ont surgi pour satisfaire la demande. Une industrie d'entreprises de données d'IA spécialisées est en train d'émerger, créant des réseaux de travailleurs sous contrat à court terme pour produire des contenus personnalisés sur commande<sup>103</sup>. Les prix sont de 1 à 2 dollars par image, entre 2 et 4 dollars par vidéo courte et entre 100 et 300 dollars par heure pour les films plus longs. Le tarif du marché pour le texte est de 0,001 dollar par mot. Les propriétaires des photos, des podcasts et des données médicales sont rémunérés à hauteur d'environ 20 à 30 % du montant total de la transaction.
- (3) **Banques de contenus** : Shutterstock, Getty, Canva ou Adobe disposent de grandes bases de media alimentées par des auteurs, rémunérés en fonction des datasets vendus. Shutterstock, leader pour les images, a vendu des licences aux fournisseurs d'IA pour plus de 100 millions de dollars, soit 12% de son chiffre d'affaires total, et devrait augmenter de 40% en 2024<sup>104</sup>. Ces licences d'images, video et musique formeront un marché adressable de 10 milliards de dollars en 2030, en croissance annuelle de 22% sur la période<sup>105</sup>.
- (4) **Groupes de media** : les propriétaires de bases de données cèdent des licences sur les datasets, sans rémunération<sup>50</sup> complémentaire pour les auteurs
- (5) **Sociétés d'auteurs** : les auteurs veulent pouvoir définir les options de leur licence pour les IA génératives<sup>106</sup>. Les idées fleurissent comme celle de *the Authors' Guild*<sup>107</sup> avec la plateforme *Created by Human*, comme *Human Native AI*<sup>108</sup>, *Fairly Trained*<sup>109</sup>, *Dataset Providers Alliance*<sup>110</sup> ou la *Commission de l'IA* en France<sup>111</sup>. À ce stade, les modèles économiques ne sont pas définis et les perspectives de développement non chiffrées.

#### Propositions de licences des œuvres pour l'IA générative

INITIATIVE	PROPOSITIONS PRINCIPALES
The Authors' Guild	<ul style="list-style-type: none"> <li>- Mise en place d'accords pour l'utilisation des œuvres</li> <li>- Paiement de royalties pour l'utilisation d'œuvres protégées</li> <li>- Transparence sur les données utilisées</li> </ul>
Created by Human	<ul style="list-style-type: none"> <li>- Label "Human-Created" pour les œuvres</li> <li>- Plateforme permettant aux auteurs d'autoriser et contrôler les licences</li> </ul>
Human Native AI	<ul style="list-style-type: none"> <li>- Place de marché d'ensembles de données de formation des IA génératives</li> <li>- Suivi des autorisations pour chaque contenu utilisé</li> </ul>
Fairly Trained	<ul style="list-style-type: none"> <li>- Label « Fairly Trained » pour les IA génératives</li> <li>- Contrats standardisés pour l'utilisation des œuvres dans l'IA</li> </ul>
Dataset Providers Alliance	<ul style="list-style-type: none"> <li>- Rémunération équitable</li> <li>- Contrats standardisés (Pay-to-train, royalties sur chiffre d'affaires, abonnement, licence par secteur)</li> <li>- Transparence sur l'utilisation des données d'entraînement</li> <li>- Autorisations individualisées par auteur</li> </ul>
Commission de l'IA	<ul style="list-style-type: none"> <li>- Plate-forme publique</li> <li>- Intermédiation de la Bibliothèque nationale de France et l'Institut national de l'audiovisuel</li> </ul>

En faisant un parallèle avec le marché de la musique, nous sommes actuellement dans la « Napster-IA », en attente des solutions équivalentes à Spotify et Apple Music. Leurs modèles économiques ont assaini le marché à partir de 2012, là où streaming et piratage étaient jusqu'alors confondus.



© Financial Times, Goldman Sachs

modèles économiques de type « sociétés d'auteurs » et leur exploration volontaire par les parties prenantes.

Malgré leur caractère embryonnaire, la multiplication de ces initiatives ouvre une nouvelle perspective de « propriété liquide ».

De même, nous pensons que le marché potentiel de valorisation des œuvres par l'IA générative est supérieur à celui des OGC, et représente une opportunité pour les ayants-droits. Symétriquement les exploitants d'IA générative pourraient accéder à des bases de contenus de qualité.

Ce potentiel, qui pour l'instant bénéficie aux segments alternatifs du marché, sera libéré par la proposition de nouveaux

## 10 – Vers la propriété liquide

### Interview de Vincent Lorphelin

Le Soir (Bruxelles), Janvier 2024



« Sam Altman et Elon Musk disent que l'IA générative va rendre le travail obsolète, il faut donc se préparer au revenu universel. Cela coagule avec la thèse de la fin du travail relayée par Jeremy Rifkin il y a quelques années. Cela rejoint aussi celle de la singularité qui avance que d'ici quelques années, l'IA générale va dépasser l'humain dans tous les domaines. La preuve c'est que l'on a toujours dit que l'IA n'ira jamais dans les métiers créatifs (elle ne sera jamais auteure, dessinateur...). Sauf que depuis l'an dernier, on y est. Les portes du bastion sont enfoncées.[...]

A l'inverse, l'ADN européen, c'est l'économie sociale de marché, c'est l'économie qui fait société. Le travail fait sens. Historiquement et d'un point de vue stratégique, la gestion collective des droits est deux fois plus développée en Europe qu'aux

Etats-Unis. De ce fait, si on en revient à cette question fondamentale de l'intérêt général, pour faire contrepoids à la thèse de la singularité, l'intérêt général européen c'est de valoriser le modèle décentralisé.

Imaginons que ce modèle de plateforme décentralisée devienne systémique, fasse système, au même titre que le font, aujourd'hui, les plateformes centralisées comme les Gafam. Dans ce cas, la matière manipulée, ce sont des droits intellectuels, pas des marchandises. C'est de la valeur économique non marchande. L'économiste Karl Polanyi avait défendu l'idée que l'économie de marché a marchandisé le travail, alors que le travail ne peut pas être assimilé à une marchandise.

Les plateformes centralisées, elles, vont tenter au maximum de donner une valeur marchande aux droits intellectuels. Dans une logique décentralisée, les droits vont s'exprimer en pourcents, pas en monnaie. Les valeurs économiques manipulées ne s'expriment plus en euros ou en dollars. C'est une économie qui est fondée non pas sur de l'argent, mais sur une propriété. Pour faire un parallèle avec ce qu'Emile Zola avait appelé l'argent liquide (avec le passage d'une économie où la richesse ne venait plus de la propriété terrienne mais de la circulation de l'argent), on propose la notion de « propriété liquide », où le fonds de roulement de l'économie c'est la circulation de droits de propriété intellectuelle.

Aujourd'hui, on a tellement tout financiarisé, qu'une entreprise présente des comptes comme si tous les jours elle était à vendre, parce qu'elle mesure tout en argent. Or, il y a des richesses immatérielles, hors-bilan, qui expriment des valeurs qui ne s'expriment pas en argent. La richesse de l'entreprise dépasse ce qu'exprime l'argent. Dès lors que l'on manipule des propriétés liquides, on est amené à trouver d'autres indicateurs. Progressivement, on évolue vers une société qui solvabilise une richesse d'un autre ordre.» (lire la suite [ici](#)).

Au même titre que le modèle de plate-forme d'Uber a inspiré le concept d'ubérisation, nous nommons « sacémisation » le phénomène de diffusion du modèle économique des sociétés d'auteur, qui valorisera le capital humain grâce à la propriété liquide.

## 11 – Promouvoir l'intelligence humaine

Brad Smith, président de Microsoft, est en train de commettre une faute morale contre la société. En parlant de l'IA générative comme ChatGPT, il dit : « Nous avons tous le droit, en vertu de la loi sur le droit d'auteur, de lire et d'apprendre. Nous nous demandons maintenant si nous pouvons permettre aux machines d'apprendre de la même manière. Je pense qu'il y a un impératif sociétal pour rendre cela possible. »

Cette intervention peut sembler anodine. Après tout, si les machines absorbent internet pour créer des articles et illustrations à un prix dérisoire, pourquoi s'en priver ? Le New York Times et d'autres ayant-

droits, qui s'opposent à cette utilisation de leur contenu, ne font-ils pas un mauvais procès à Microsoft ? Si les machines ne font pas de plagiat, où est le mal ? Celui-ci est dans la finalité inacceptable de ce discours.



### Un flou judiciaire avantageux

On le sait, les plateformes des Gafam moissonnent une masse de petites contributions et promeuvent le travail gratuit. Elles changent encore de dimension avec l'IA, dont la voracité nécessite la conquête de nouveaux territoires, ceux des auteurs.

Par exception, le législateur avait défini le « fair use » pour autoriser Google à publier des extraits de contenus privés. Comme le résultat judiciaire est encore incertain, l'ancien avocat Brad Smith sait que l'opinion publique jouera un grand rôle. C'est pourquoi il distille l'image mentale frelatée

du « droit aux machines d'apprendre ». Son « impératif sociétal » n'est qu'une vulgaire contestation de la légitimité des auteurs à défendre leurs territoires.

Il détourne l'attention en débattant de questions faussement éthiques. L'intelligence humaine est-elle condamnée à être déclassée ? Peut-on laisser l'IA prendre des décisions ? Il a beau adopter la posture du sage face à un Elon Musk moins policé, ils sont tous les deux les augures de la même illusion : celle du dépassement de l'intelligence humaine.

## **IA : une illusion de supériorité ?**

Le meilleur roman ne remplacera jamais l'expérience de la vie. La meilleure «visio» ne se substituera jamais à une vraie rencontre. Aussi perfectionné soit-il, le modèle n'est pas le réel. Chacun de nous a une intelligence du monde qui dépassera toujours celle des machines, qui n'ont d'intelligence que celle de ses modèles.

Les IA éclairent notre intelligence . Elles permettent à la police de l'environnement de détecter les pollueurs, à l'enseignant de comprendre les difficultés de mille élèves, ou au laboratoire d'anticiper les effets d'un médicament.

Répandre l'illusion de la supériorité prochaine des machines est irresponsable. C'est nier le travail strictement humain, qui puise justement son inspiration entre le monde et ses modèles, lorsque l'artiste révèle un non-dit, lorsque le conducteur compense des manettes qui répondent mal, lorsque le médecin doute d'un diagnostic trop évident, ou lorsque le juge se fait une conviction intime.

### **La défense des auteurs**

Nourrir l'illusion d'une compétition entre intelligences et attribuer des droits humains aux machines est une faute contre la société. Il faut affirmer au contraire la prééminence de l'intelligence humaine et garantir l'intégrité de ses territoires.

L'illusion de Brad Smith n'aurait pas de gravité si elle n'était le fait du leader mondial de l'IA. Gagnée par cette illusion, la Commission Interministérielle de l'IA affirme que les machines dépasseront les humains. Elle défend, heureusement, la juste rémunération de la création humaine, mais sans se rendre compte que c'est l'appréciation même de cette justesse qui est manipulée.

L'Europe a déjà décidé avec l'AI ACT de protéger les droits d'auteur. Elle doit maintenant établir le cadastre des territoires exploités par les IA et défendre leur loyer. ([Lire](#) la suite dans Les Echos du 6 mai 2024)



## 12 – Une solution sacémisante pour l'Europe

### Entretien avec Jean-Paul Betbèze

Economiste, Membre du Comité scientifique de la Fondation Robert Schuman



### Quels devraient être les points clés de la “sacémisation“ ?

JPB : « Une solution “sacémisante“ pour l'Europe devrait :

- Être fondée sur la propriété intellectuelle des auteurs, comme les OGC. Il faut maintenir et garantir la propriété intellectuelle des auteurs, qui pourrait, sinon, être dissoute ou disparaître dans la révolution actuelle de la communication. La communication se mondialise, c'est la base du changement majeur que nous vivons, avec la multiplication des messages en tous sens. L'apport de chacun doit être reconnu, sauf à assécher le mouvement mondial en cours par une concentration excessive au bénéfice de certains auteurs, les autres ne trouvant aucun avantage à leur travail et à leurs idées, aucune reconnaissance, aucune rémunération. À cause des écarts de disponibilité des capitaux, à l'avantage des Etats-Unis pour une grande part, le danger

est une monopolisation des sources apparentes de création.

- Collecter des redevances sur l'exploitation des œuvres par l'IA. La “sacémisation“ permet non seulement de tracer, mais aussi d'identifier et estimer les sources, même lorsqu'elles semblent minimes. Ce processus permet de conserver toutes les origines et donc toutes les reconnaissances de créations et d'améliorations : rien ne se perd, tout se garde.
- Reconnaître la créativité et la distinguer du contenu amateur. Dans la prolifération des travaux, il importe de pouvoir mettre en avant ceux qui sont originaux, pour les signaler et les valoriser.
- Mettre en œuvre des mécanismes d'incitation individuelle, pour que le prix de la créativité soit compétitif sur le marché et compatible avec le coût de production. Le risque de la multiplication des travaux est la banalisation, d'autant plus que leur coût de production physique est évidemment faible. Une facturation liée au seul coût de production court le risque d'être désincitative. Le grand avantage de la “sacémisation“ est de s'éloigner d'une logique de prix liée au coût de production, pour aller vers une autre logique, celle des marchés, qui est fondée sur l'importance du succès, la répétition du message.

### Quel est l'avantage d'un marché dynamisé par la “sacémisation“ ?

Dans cette logique de marché, c'est la demande mesurée, d'où découle la demande anticipée, qui fabrique les nouveaux prix, donc les marges. Dans ce contexte les marges ne peuvent que croître, donc les profits

marginiaux, donc les valorisations. Cette différence entre les coûts faibles, dépistés par les moteurs de recherche et les valorisations qui auront ainsi une base objective, mises à jour par les outils de type Sacem, est la base de la logique proposée. Elle permettra de mettre en avant les apports de chacun, notamment dans un contexte européen.

### Et pour les activités créatives ?

N'oublions pas que la concurrence monopolistique est toujours à l'œuvre : elle pousse à la réduction du nombre de concurrents, jusqu'à ce qu'un seul émerge : c'est le monopole. Là plus qu'ailleurs, le gagnant rafle toute la mise. La "sacémisation" doit s'y opposer en permettant des rémunérations mieux réparties avant qu'un seul monopoleur n'apparaisse. Ceci est évidemment favorable à la recherche, à l'innovation, à la création, puisque l'on sait bien que la situation de monopole se retourne contre le monopoleur lui-même, dans une situation de marché qui s'assèche à terme, car sa dynamique conduit à la baisse de ses profits. La "sacémisation" respecte ainsi les créations et freine la sélection impitoyable et, au fond, destructrice du monopole lui-même.

Dans le marché des idées, la vitesse est mortelle, la mémoire des apports est décisive, en liaison avec leur juste rémunération. C'est aujourd'hui possible et, on le comprend, plus nécessaire que jamais. Il s'agit de sauver la création grâce aux progrès qu'offre la technologie. »

Michelle Bergadaà, Professeure émérite à l'université de Genève.

Co-auteurice de *Réinventer l'intégrité académique à l'ère de l'intelligence artificielle* (à paraître, EMS, février 2025)



### Vous dites que l'AI Act ne concerne pas que les auteurs, mais aussi les chercheurs, pourquoi ?

MB : « Nous devons réfléchir à appliquer l'AI Act pour rémunérer les auteurs de « recherches authentiques ». La France paye mal ses chercheurs, cela freinerait leur fuite vers l'étranger ou le privé. Cela coûte bien plus cher à la société d'écraser sous la bureaucratie ses jeunes chercheurs que de bien les payer. Beaucoup se démotivent et font des heures supplémentaires d'enseignement au lieu de faire de la recherche. Alors si l'Etat ne peut pas les rémunérer convenablement, il faut bien que le système de production des articles, livres, rapports le fasse. Cela doit être la pierre angulaire de la stratégie européenne de l'IA générative. Avec en prime, un système automatisé de détection des fraudes.

### La fraude n'a-t-elle pas toujours existé ?

La crise des opioïdes aux USA [Le fentanyl et apparentés, prescrits par ordonnance, ont à eux seuls été responsables d'environ 71.000 décès par overdose de drogue en 2021] est arrivée par un seul et unique article faux, mal cité et repris. C'était bien avant l'IA. Avec l'IA, la diffusion des articles frauduleux se fait essentiellement via de nouveaux articles. C'est un nouveau modèle d'essaimage du faux.

Aujourd'hui trois temporalités cohabitent. Premièrement le temps instantané et émotionnel (de Trump et de la génération Z) permet de produire de manière non seulement des fausses informations, mais aussi de faux auteurs, tout aussi crédibles. Deuxièmement l'éternel présent des communautés de jadis entretient un monde alternatif où la terre est plate et où la javel élimine la Covid. Troisièmement le temps linéaire est poussé par la technologie matérialiste d'Elon Musk. L'IA est un accélérateur de fraude qui foisonne dans les interstices de ces temporalités.

### **Comment agir ?**

Pubpeer et Retraction Watch Database cherchent à coincer des délinquants individuels, un par un. Mais trois ans sont nécessaires pour boucler l'enquête et obtenir la rétractation dans les journaux.

Il faut appréhender ce phénomène de manière systématique avec des radars d'intégrité, une injonction pour les fraudeurs de rendre l'argent perçu, et une exclusion des bases de données de l'IA. Le futur de notre société, c'est les jeunes que nous formons. Aidons-les par notre posture ferme et en rémunérant leurs créations originales par l'IA générative. Cela motiverait bien des jeunes chercheurs intègres à créer de la connaissance. »



# Recommendations

## 13 – La solution Controv3rse

Sam Altman, fondateur d'Open AI, souhaite disposer « de nouveaux modèles économiques qui permettront aux créateurs de générer des revenus. »

Nous avons vu que le modèle de l'Uber des données émerge pour éviter la dégénérescence de l'IA générative, mais qu'il menace en revanche le capital humain. Le modèle économique doit s'appuyer sur les OGC pour ouvrir une perspective de revalorisation de ce capital par la propriété liquide et la sacémisation.

Le modèle économique que nous recommandons à Sam Altman, et aux autres exploitants d'IA générative, est le suivant :

### Solution Controv3rse

RECOMMANDATION	OPEN AI	OGC
1 - Réserve	Réserver 15% du chiffre d'affaires des IA génératives pour la rémunération équitable des auteurs et ayant-droits	Mettre à disposition des auteurs une plate-forme pour s'opposer, le cas échéant, à l'utilisation de leurs œuvres par l'IA générative, ou pour choisir une licence, œuvre par œuvre
2 – Segmentation	Identifier le principal OGC par segment de marché (media/dataset * pays)	Revendiquer un segment de marché (media/dataset * pays)
3 - Méthode	Publier la méthode d'attribution par OGC, détailler la désambiguïsation en cas de recouvrement	Négocier des accords de coordination ou de réciprocité entre OGC <sup>112</sup>
4 – Non-Attribution	Identifier un fond par pays pour les montants non attribués <sup>113</sup>	Revendiquer la gestion du fond
5 - Rapports	Publier le résumé suffisamment détaillé <sup>114</sup>  Adresser à chaque OGC identifié un rapport mensuel contenant les identifiants des auteurs, le détail de l'utilisation de leurs œuvres, par contenu synthétique, et la valeur attribuée	Adresser à l'exploitant la liste des œuvres non utilisables par l'IA générative
6 - Paiement	Répartir la réserve aux OGC et aux fonds	Répartir les montants collectés aux auteurs et ayant-droits

# 14 – Enclencher la négociation

## Étapes

Pour aboutir à des accords entre exploitants d'IA générative et OGC, nous recommandons aux parties prenantes de documenter les étapes suivantes :

### Étapes de la solution Controv3rse

RECOMMANDATION	PROCESSUS	BONNES PRATIQUES, COMPARABLES
1 - Prototype	Discuter à partir d'une solution technique à parfaire pour faciliter la projection des interlocuteurs dans un résultat, et réduire la complexité du débat <sup>115</sup> (voir notre démonstrateur)	Démonstrateur, méthodes essai-erreur
2 - Vision	Développer la prospective : décentralisation vs. plateformes centralisées, sacémisation vs. IA ubérisante, travail des auteurs vs. marchandisation, absurdité du « droit » des machines, propriété liquide, métriques de l'utilité	Cadre européen : économie sociale de marché, économie de la créativité, héritage et diversité culturelle
3 - Rémunération équitable	Préciser le taux de rémunération équitable <sup>116</sup> à partir de la mesure de l'utilité de l'ensemble des œuvres d'auteurs pour l'ensemble des contenus générés. Nous proposons une valeur de départ de 15%.  Proposer les indicateurs d'utilité des contributions pour une attribution auteur par auteur, et la méthode d'affinage de cette attribution par anticipation des futurs progrès techniques.	Spré Directive 2006/115/CE (13) <sup>117</sup> Droits voisins (73) <sup>118</sup>  Statistiques Yacast pour la Spré
4 – Code des bonnes pratiques par le Bureau de l'IA	Observatoire de l'état de l'art <ul style="list-style-type: none"> <li>- Méthodes : valeur de Shapley, pay-to-train, similarité</li> <li>- Arbitrages : paramètres, granularité, indicateurs d'utilité</li> <li>- Plate-formes d'option et de licences</li> <li>- Identification des ayant-droits et auteurs</li> <li>- Traitement des montants non attribués</li> </ul> Résumé suffisamment détaillé : <ul style="list-style-type: none"> <li>- Contenus</li> <li>- Granularité<sup>119</sup></li> <li>- Format</li> </ul> Rapports aux ayant-droits : <ul style="list-style-type: none"> <li>- Revendications des datasets par les ayant-droits</li> <li>- Méthodes d'attribution et de répartition par ayant-droit</li> <li>- Traitement des recouvrements</li> <li>- Données</li> </ul> Contenu secret <sup>120</sup> : Informations non divulguées dans les résumé et rapports  Normalisation progressive	Code des bonnes pratiques contre la désinformation <sup>121</sup>  Recommandations de l'Authors' Guild <sup>122</sup>  Mission du CSPLA <sup>123</sup>  ISWC <sup>124</sup> , ISAN <sup>125</sup>
5 – Brevets	Assurer les positions française et européenne par un portefeuille de brevets dédiés, en coopération publique-privée	Programme Alliance PI de l'INPI <sup>126</sup>

## Rémunération équitable

La rémunération équitable par l'IA générative pourrait s'inspirer de celle de la Spré. Cet organisme prélève en France une redevance calculée sur le chiffre d'affaires des discothèques au bénéfice des musiciens. Un échantillon représentatif de 120 d'entre elles est équipé d'un boîtier « Yacast », financé par la Sacem et la Spré. Yacast collecte les listes de musiques diffusées, et publie des statistiques<sup>127</sup> par artiste, label, genre et nombre de semaines de diffusion. Après retenue statutaire de 9%, la Spré répartit le prélèvement aux OGC (SCPA 50%, ADAMI 25% et SPEDIDAM 25%)<sup>128</sup> et leur communique les statistiques Yacast pour rémunérer les ultimes bénéficiaires (artistes interprètes, producteurs, actions d'intérêt général).

## Taux équitable

Dans son rapport de 2020, *L'auteur et l'acte de création*, Bruno Racine déplorait que « la rémunération proportionnelle des auteurs, n'atteint ou dépasse qu'exceptionnellement 10 % des recettes d'exploitation »<sup>129</sup>. Il attribuait ce constat à « la relation qui lie l'artiste-auteur aux acteurs de l'aval (éditeurs, diffuseurs, producteurs) [qui] apparaît profondément déséquilibrée, ce qui conduit, le concernant, à mettre en cause dans de nombreux cas l'idée même de liberté contractuelle »<sup>130</sup>. Pour y remédier, il voulait « renforcer les artistes-auteurs collectivement, par l'organisation rapide d'élections professionnelles qui permettraient de donner corps et légitimité au Conseil national des artistes-auteurs à créer afin de servir de cadre à la négociation collective avec les diffuseurs » pour fonder « un dispositif contractuel rémunérant le temps de travail de l'auteur, préempté par le commanditaire ».

Comme l'idée de revaloriser la valeur-travail n'est pas applicable au cas de l'IA générative, nous proposons de définir le taux équitable comme la valeur de l'utilité de l'ensemble des contributions des auteurs rapportée au chiffre d'affaires des contenus générés<sup>131</sup>. C'est sur ce principe qu'est conçue la valeur de Shapley et que l'économiste Ernst Fehr évalue à 14% la valeur apportée par la presse à Google Search<sup>132</sup>. Depuis la fameuse question « ce commentaire vous a-t-il été utile ? » lancée par Amazon<sup>133</sup>, tous les sites web commerciaux améliorent la qualité de leur offre en multipliant les métriques de l'utilité, qui sont maintenant d'usage courant<sup>134</sup>.

Fixer l'assiette sur le chiffre d'affaires du contenu généré permettrait à la rémunération équitable de l'auteur de ne pas dépendre de la plus ou moins grande segmentation verticale de la filière, par exemple : auteur / plateforme / dataset / modèle général / fine tuning / interface / génération.

## 15 – Licence -Remerciements

Cette note et ses tableaux sont accessibles sous licence CC-BY, sous réserve du partage équitable des fruits de son exploitation par IA générative, concédée par l'association Controv3rse, RNA W751267044.

Elle s'inspire des réflexions et échanges entre les membres des missions ministérielles pour l'AI Act, ceux des think tanks indépendants Controv3rse et Institut de l'Économie, ainsi que d'autres personnalités reconnues pour leur expertise. Elle prolonge l'étude<sup>135</sup> faite précédemment pour la mission parlementaire Pradal-Rambaud<sup>136</sup> et des tribunes publiées dans *Le Monde* et *Les Echos*. Vincent Lorphelin, Fondateur et co-Président de ces think tanks, rédacteur de cette note, les remercie chaleureusement pour leurs idées et la qualité de ces échanges.

## Présidents et Rapporteurs des missions parlementaire et ministérielle :

Alexandra Bensamoun, professeure de droit à l'université Paris Saclay, ancienne membre de la Commission de l'intelligence artificielle

Joëlle Farchy, professeure des universités, Directrice du m2ecn et de la Chaire PcEn  
Université Paris 1 Panthéon-Sorbonne, membre du CSPLA

Philippe Pradal, Député de la Troisième Circonscription des Alpes Maritimes, Assemblée Nationale

Stéphane Rambaud, Député de la Troisième Circonscription du Var, Assemblée Nationale

Lionel Ferreira, Maître des requêtes au Conseil d'Etat

Julie Groffe-Charrier, Maître de conférences HDR en droit privé chez Faculté Jean-Monnet

Bastien Blain, Professeur Junior en économie, scientifique des données et du comportement,  
Université Paris 1 Panthéon-Sorbonne

## Experts et Entrepreneurs :

William Bailey  
co-Fondateur et président, Bolero Music

Emmanuel Benazera  
Président Directeur Général, Jolibrain

Jean-Paul Betbèze  
Economiste, Betbèze Conseil

Michelle Bergadaà  
Présidente à l'Institut de Recherche et  
d'Action sur la Fraude et le Plagiat  
Académiques

Michel Bokobza  
Chargé de missions, Collège de Paris

Laurent Cervoni  
Directeur de la Recherche et de  
l'Innovation, Talan

Charles-Éric de la Chapelle  
Fondateur, Myriad Data

Jean-Philippe Clair  
Directeur Marketing, Communication et  
Innovation, Keyrus

Alice Coatalem  
co-Fondatrice de CogNeed et Professeur  
Associée, Université Paris Dauphine PSL

Christophe Collet  
Fondateur, AskLocala

Frédéric Dayan  
Fondateur, Exactcure

Guillaume Desveaux  
co-Fondateur d'Aleia et Administrateur,  
AI Cargo Foundation

Jonathan Dory  
Fondateur, Live Crew

William Eldin  
Fondateur, XXII

Stéphanie Flacher  
co-Fondatrice, Logion Network

Raphaël Frisch  
co-Fondateur, HawAI.tech

Alain Garnier  
CEO, Jamespot

Anthony Graveline  
Fondateur, Disaitek

Philippe Guguen  
Président, Sorga - Map Emulsion

Francis Hachem  
Fondateur, Codenekt

Rodolphe Hasselvander  
Fondateur, Blue Frog Robotics

Brice Hoarau  
Fondateur, Semdee

Matthias Houllier  
co-Fondateur, Wintics

Casey Joly  
Avocate, IPso

Jean Latger  
Fondateur, Oktal-SE

Frédéric Lefebvre-Naré  
Directeur data IA, Niji

Youness Lemrabet  
Fondateur, Everysens

Jacques Lévy-Vehel  
Président de Case Law Analytics, et  
Directeur de Recherches, INRIA

Olivier Laborde  
Leader Innovation et Transformation  
digitale, BPCE

Alexandre Leforestier  
Fondateur, Panodyssey

Sixtine Lorphelin  
Ingénieure IA, UTC, étudiante à  
l'INSEAD

Aymeric Masurelle  
co-Fondateur, Spoon.ai

Pierre Miralles  
co-Fondateur, Footovision

Emmanuel Moyrand  
co-Fondateur, France Meta

Clément Merville  
Fondateur, Manzalab - Teemew

Edouard de Miollis  
Fondateur, Polycube

Nathalie Nevejans  
Professeure de droit privé et intelligence  
artificielle, Université d'Artois

Rémy Ozcan  
co-Fondateur, Crypto4All

Jean-Jacques Quisquater  
Professeur de Cryptographie,  
Polytechnique Louvain, chroniqueur du  
Monde

Jean-Michel Salomon  
Président, Société des Auteurs de Jeux

Frédéric Soufflet  
co-Fondateur, Haapie

Clément Tequi  
co-Fondateur, Ternoia

François-Xavier Thoorens  
CEO, Vaultys

Arnaud Touati  
Avocat Associé, #Hashtag

Christophe Tricot  
co-Fondateur, La Forge


Killian Vermersch  
co-Fondateur et Directeur Général,  
Golem.ai



# The Controv3rse solution

An economic model to enhance cultural and human capital through generative AI



 Adobe Firefly

C O N T R O V **3** R S E

# Summary



Controv3rse is an independent think tank, formed by 70 entrepreneurs and experts in generative AI. This report proposes a solution in response to the statement by Sam Altman, founder of OpenAI, who advocates for a new economic model rewarding creators in the face of the rise of generative AI.

It must be said that the conflict between authors, whose works are being plundered, and the exploiters of generative AI, who need clarity and legal stability, will not be resolved in court for at least 10 years.

In the meantime, the proliferation of generative AI dilutes originality and erases traces of source works. In parallel, micro-tasking platforms respond to the massive demand for data to the detriment of models based on creativity and copyright, such as SACEM. Cultural and human capital, once central to European strategy, is doubly threatened with extraction.

However, Nobel Prize-winning economist Jean Tirole emphasizes that national wealth is based on the valorization of innovation, supported by intellectual property as a lever for creativity. To preserve cultural and human capital while stimulating innovation, the new economic model must reinvent that of authors' societies.

Lloyd Shapley introduced in 1953 a key concept for fairly distributing gains in a cooperative game, based on the contribution of each player. The derived method, ideal for remunerating the authors of generative AI, unfortunately has a prohibitive computational cost. Alternatives exist, such as the "pay-to-train" model used by image banks (Shutterstock, Getty). They remunerate authors according to the quantity of content provided, but ignore the quality or usefulness of the works. Startups and researchers are completing this model to offer a fairer distribution based on similarity. We offer the reader an online simulator that will allow him to see for himself, in real life, the technical and economic feasibility of this approach.

In perspective, this model protects the rights of authors and opposes "Uberizing AI". "Sacemization" values "liquid property", the intellectual wealth of creativity, which maintains and develops cultural and human capital in the face of its extraction by monopolies.

We therefore recommend to Sam Altman and the operators of generative AI an economic model in cooperation with the authors' societies, and to all stakeholders a method for initiating negotiations in this new framework of reflection.

# Challenges

# 1 – Who wants generative AI to pay authors?

“I think we need a new agreement, a new standard, a new protocol, whatever you want to call it, for how creators are going to be rewarded,” said Sam Altman, founder of Open AI. “We need to have new business models that are going to allow creators to generate revenue.”

The idea of paying authors through gen-AI has been advocated by other personalities for several months.

## AI operators who want to pay authors

NAME	COMPANY	QUOTE
Sam Altman	Open AI	“I think we need a new agreement, a new standard, a new protocol, whatever you want to call it, to determine how creators are going to be rewarded. We need to have new business models that are going to allow creators to generate revenue.”
Arthur Mensch	Mistral AI	"We have an interest in having access to quality content. We have therefore started discussions with content providers, in publishing and in the press. There are synergies and value sharing mechanisms to be found."
David Holz	Midjourney	“There’s no real way to take a hundred million images and figure out where they came from. It would be nice if the images had metadata embedded in them about who the copyright owner was or something. But that’s not possible, there’s no registry. There’s no way to find an image on the internet and then automatically trace it back to its owner and then be able to do anything to authenticate it.”
Chris Dixon	a16z	"The [Web 3] movement has the potential to bring back the spirit of the early Internet, secure creators' property rights, reclaim user ownership and control, and break the stranglehold of large, centralized corporations on our lives."
Neal Stephenson	Laminal (inventor of metaverse)	“You can actually track how much influence a given image or input had on the end result. And then, if that end result is worth something, then maybe there’s a way to give credit and proportional compensation for those inputs [...] Let’s say I was writing a book that had a magic sword that was just described in the book. [...] Someone could then create an asset that could be sold on the Unreal Asset Store [...] You have sound designers who have to do something similar with the sounds that it makes when it’s used. And you have programmers who have to, using blueprints or C++, integrate the sword into the game so that it’s actually able to do something and contribute to the experience. So at each step, more value is added. And at the end of that process, you have something that could actually generate revenue. And when that revenue finally shows up, what you want it to do is propagate backwards.” And you would like the different people who contributed to the value chain to be compensated in some way"
Jaron Lanier	Microsoft	“People will be paid for their data and will pay for services that require data from other people. Individuals’ attention will be guided by their self-defined interests rather than by manipulative platforms beholden to advertisers or other third parties. Platforms will be given higher-quality data with which to train their machine learning systems.” Lanier acknowledges that even data dignity researchers can’t agree on how to untangle everything that AI models have absorbed or how much detail is needed to account for it. Still, Lanier thinks it could be done—incrementally. “We need to calculate and present the provenance of the human sources that were most important for a given AI outcome. We don’t do that now. We can do it effectively and efficiently, we just don’t. It has to be a societal decision to take that step.”

NAME	COMPANY	QUOTE
Yat Siu	Animoca Brands	“Data is the most valuable resource. For starters, it powers all the artificial intelligence we hear so much about. Without data, there’s no ChatGPT . Data is the new job. And we’re not getting paid fairly for it. [...] There’s a clear correlation between property rights and the wealth of nations.” “Imagine going on Facebook and at the end of the day realizing you just made a thousand dollars for the company. Your relationship with Facebook would change completely. You’d probably demand your fair share.”
Yann Le Cun	Meta	"a model of propagation of [copyright] rights in one direction and of revenues in the other"
Trip Adler	former CEO of Scribd	“The Fourth Law is a set of guiding principles for how AI companies can use and train on human-created content. The Fourth Law, inspired by science fiction author Isaac Asimov’s Three Laws of Robots, states that humans should have the right to consent to and control how AI uses their work, to be compensated (if requested), and to be credited for their work (if a book is referenced in the output, there should be a link to purchase it).”

Given Sam Altman's leadership, the idea of compensating authors takes on more substance, although he considers that this compensation would only be due when the author's style is plagiarized, which is very limiting.

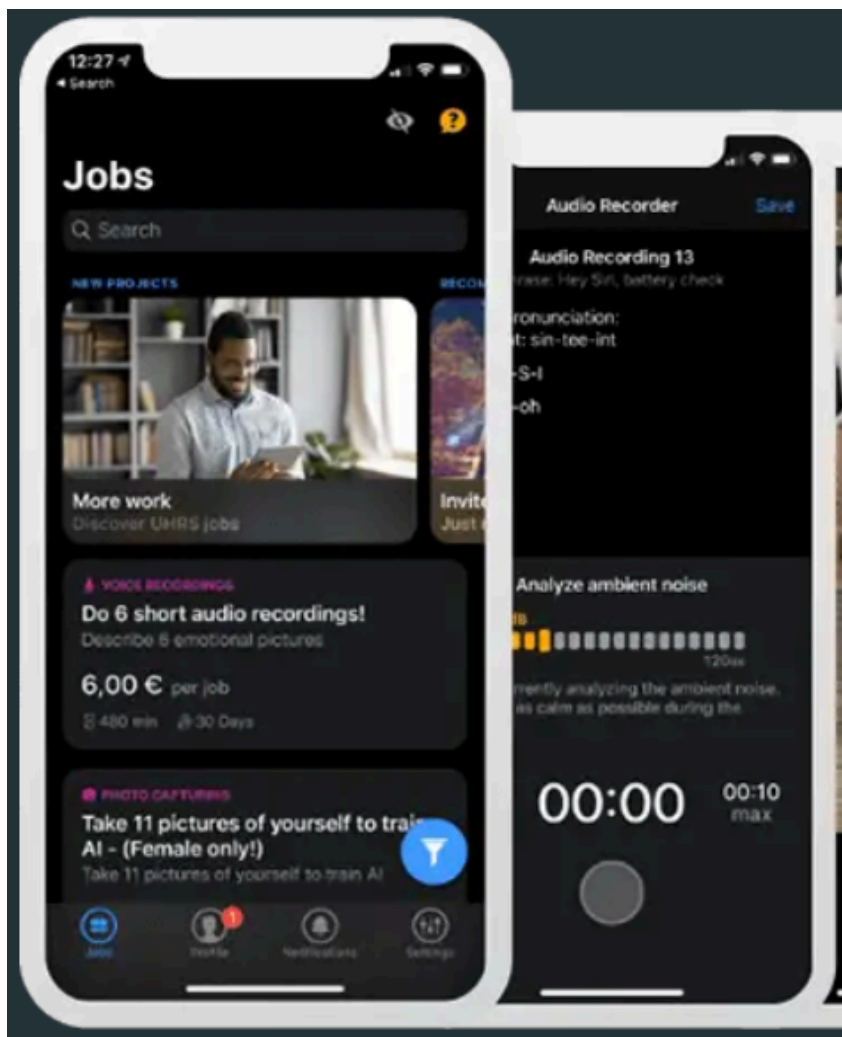
However, he questions the economic model of its implementation . The purpose of this note is to propose this new economic model within the framework of the AI Act .

## 2 – The three faces of authors’ remuneration

If dealt with on a purely legal level, the conflict over scraping , which pits authors against operators of generative AI, will last at least ten years.

In the meantime, the sector will segment into data creators, platforms, aggregators, negotiators, trainers, refiners, exploiters and integrators of generative AI, which will dilute the act of creation. Cultural heritage will have been mixed, disseminated and recycled in content whose original traces will have been erased.

For fear of a degeneration of learning content, a new sector of micro-tasks is already developing for the mass production of content.



©Clickworker 2024

economy platforms for AI are based on command logic. Clickworker pays users €6 directly to describe 6 images out loud and post the recording. Or to take a picture of themselves doing physical exercises. Gigwalk and Premise pay \$4 to take a picture of a traffic jam at an intersection. The result is used to train AIs.

This “Uber of data” type model is growing rapidly thanks to the massive need for data from generative AI operators.

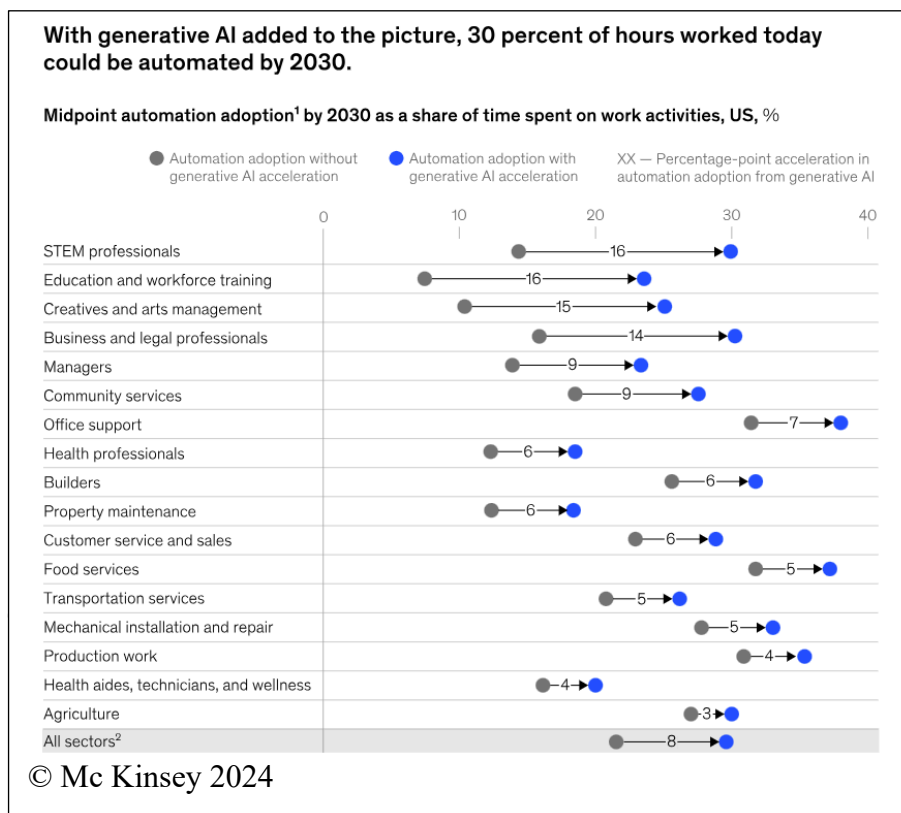
This is the opposite of copyright societies such as SACEM, based on the supply and intellectual property of authors, which track this property, collect royalties, and remunerate rights holders according to predefined distribution scales.

The first model favors demand, execution, immediate commodification,

cognitive neo-Taylorism, standardization of work. The second favors supply, initiative, intellectual property, creativity, cultural diversity. The first exploits cultural capital, the second cultivates it.

Among other economic models (see below), scraping, data Uber and authors' societies are the three main faces of remuneration – or not – of authors, the imbalance of which aggravates current tensions.

### 3 – Beyond copyright, human capital



Faced with the development of generative AI, authors and creativity are on the front line. Next will be crafts, artistic professions, expertise, research and all professions with know-how, in all economic sectors.

As the “Uber of data” model becomes more widespread, it will give rise to a form of Uberizing AI, in the same way that the Uber company once initiated the phenomenon of Uberization.

For example, customizing the enterprise generative AI with all its data archives makes it possible to obtain a draft for each new message, report, project, presentation, pricing. The corrections of the professionals train the AI until the error rate becomes

acceptable. The productivity gains are significant but, in return, the positions of the company on the strategic links of value creation decrease to the benefit of the operator of the AI.

Since all these skills are the intersection of knowledge, practices, and social organization materials, it is the fabric of companies that is at stake. It is as if, at the time of the industrial revolution, a cartel of steam engine owners controlled the Taylorist organization of work. This is what the stock market anticipates for Big Tech. By granting capitalizations in the trillions, it no longer values companies, but monopolies.

The idea that machines have the “right to learn” for free will spread. In ten years, even if the authors end up winning their legal battle against scraping, their compensation will be based on the comparables provided by the Ubers of data, which is to say, not much. In retrospect, we will realize that existing copyright law will have been nothing more than an easy-to-circumvent Maginot Line.

The economy has cultural and human capital, long identified as a considerable asset, to the point of having been placed at the center of the European economic strategy. This strategy, too early, was finally forgotten due to the lack of an economic model. Undervalued and poorly defended, this capital is now threatened with extraction by Uberizing AI. For each company, country or continent, refusing technology is condemning itself to a lasting decline, like that of China for two centuries. Accepting it without compensation is allowing itself to be colonized without a fight.

Nobel Prize winner Jean Tirole teaches us that 1/ the wealth of nations depends on their ability to capture value at the innovation level. 2/ innovation comes from individual initiative, because the essence of creative work is that we don't know what we're going to find. 3/ intellectual property is a necessary evil to stimulate

this creativity. 4/ preventing a company from marketing its products until it has paid its royalties is a very effective weapon against Gafam . 5/ intellectual property pools with individual licensing authorization would force them to acquire licenses, but at a competitive price.

Nations that want to enhance their human and cultural capital must find a modern sacemizing solution , whose economic model will effectively compete with Uberizing AI .

# Technical solution



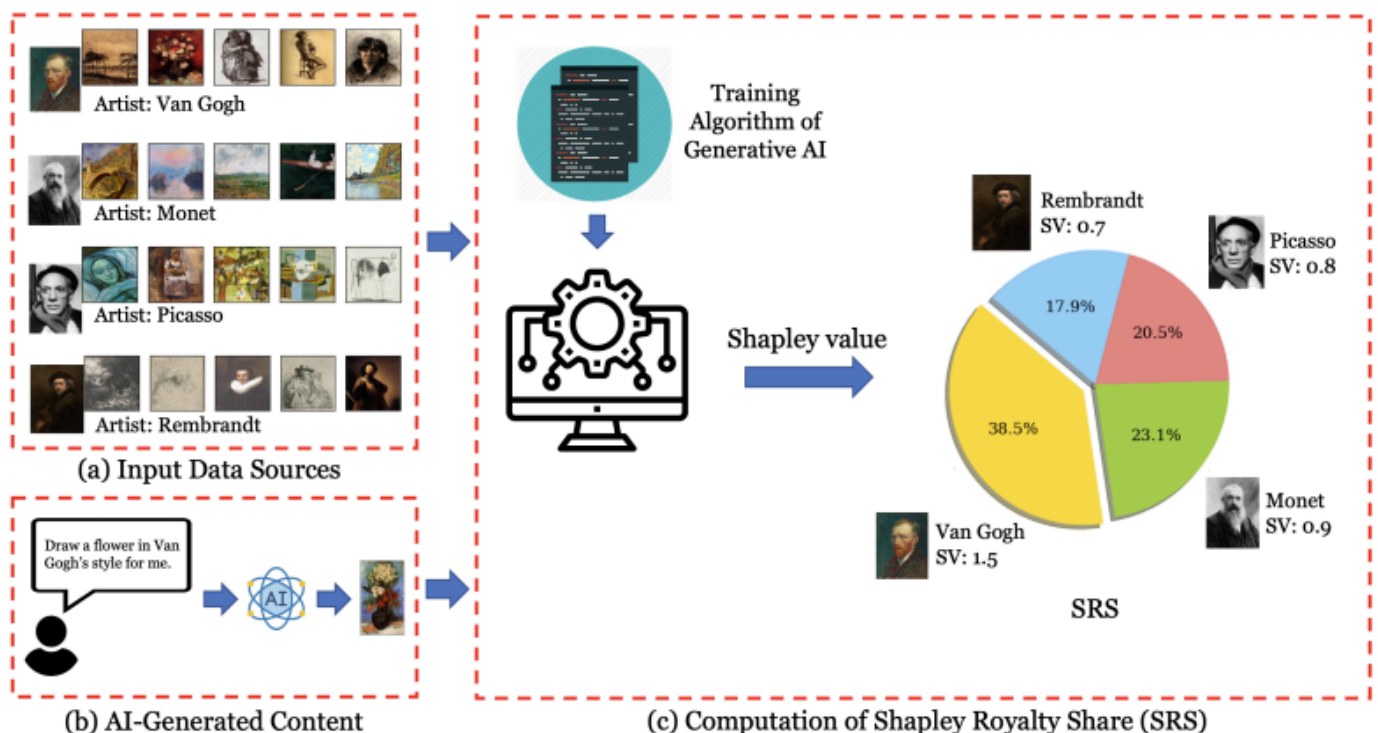
## 4 - Ideal solution

In game theory, Lloyd Shapley introduced the idea in 1953 that it was possible, in a cooperative game, to calculate a fair distribution of gains between players. For example, consider a marine ecosystem with all its living species - algae, anemones, shells, starfish - and measure the utility of each species by its contribution to life in the ecosystem. When the starfish are removed, 50% of the other species disappear, compared to 10% when the anemones are removed. The Shapley value of starfish is five times higher than that of anemones.

The ideal solution for remunerating authors is a consensus among researchers. It involves calculating the Shapley value for each author and distributing remuneration pro rata to this value.

In the case of generative AI, the fair distribution of training artwork remuneration should ideally be based on the usefulness of the training artwork for the AI-generated content, by author. In the figure below, Van Gogh's Shapley value is 1.5, or 38.5% of the total remuneration.

Shapley value



In practice, the cost of calculating the Shapley value is prohibitive, which raises the question of the technical feasibility of a fair remuneration solution.

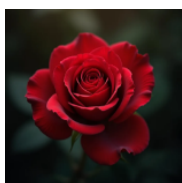
## 5 – Demonstrator

### Object

The reader will find an online simulator at [https://huggingface.co/spaces/TheFrenchDemos/genai-rew\\_a\\_rd](https://huggingface.co/spaces/TheFrenchDemos/genai-rew_a_rd)








We believe that it is now technically and economically possible to calculate an attribution score for training content in order to fairly distribute remuneration between the rights holders of this content. This simulator will allow everyone to verify their own economic hypotheses for themselves.

For example, an image of a rose was generated with the prompt "*A vibrant red rose in full bloom, macro photograph*":



The simulator identifies in a few seconds, among the ten million contents of the Open Images database, the eight original images that have contributed the most to the formation of the generated content. It calculates an attribution rate, a remuneration, and provides identification information of the original content and its author.

#### Attributions of rose images

CONTENT	ATTRIBUTION	CONTENT REWARD (c€)	CONTENT URL	AUTHOR NAME	AUTHOR URL
	21%	0.63	<a href="https://www.flickr.com/photos/50093642@N03/4629227563">https://www.flickr.com/photos/50093642@N03/4629227563</a>	Yoko Nekonomania	<a href="https://www.flickr.com/photos/nekonomania/">https://www.flickr.com/photos/nekonomania/</a>
	18%	0.54	<a href="https://www.flickr.com/photos/22305657@N02/11733854773tt">https://www.flickr.com/photos/22305657@N02/11733854773tt</a>	Bruno Pinho	<a href="https://www.flickr.com/photos/brunopinho/">https://www.flickr.com/photos/brunopinho/</a>
	15%	0.45	<a href="https://www.flickr.com/photos/12692992@N00/170278141">https://www.flickr.com/photos/12692992@N00/170278141</a>	Dean + Barb	<a href="https://www.flickr.com/photos/barbanddean/">https://www.flickr.com/photos/barbanddean/</a>
	12%	0.36	<a href="https://www.flickr.com/photos/33140999@N07/3509399376">https://www.flickr.com/photos/33140999@N07/3509399376</a>	Michael Arrighi	<a href="https://www.flickr.com/photos/arrighi/">https://www.flickr.com/photos/arrighi/</a>
	10%	0.30	<a href="https://www.flickr.com/photos/55658968@N00/8730223835">https://www.flickr.com/photos/55658968@N00/8730223835</a>	Toshihiro Gamo	<a href="https://www.flickr.com/photos/dakiny/">https://www.flickr.com/photos/dakiny/</a>
	9%	0.27	<a href="https://www.flickr.com/photos/34226407@N04/8734050294">https://www.flickr.com/photos/34226407@N04/8734050294</a>	Edd Prince	<a href="https://www.flickr.com/photos/eddpics/">https://www.flickr.com/photos/eddpics/</a>
	9%	0.27	<a href="https://www.flickr.com/photos/41353201@N07/4267426012">https://www.flickr.com/photos/41353201@N07/4267426012</a>	Musicman Loxton	<a href="https://www.flickr.com/photos/dj-dwayne/">https://www.flickr.com/photos/dj-dwayne/</a>

	8%	0.24	<a href="https://www.flickr.com/photos/64574638@N04/6006728032">https://www.flickr.com/photos/64574638@N04/6006728032</a>	serbianstruhpro	<a href="https://www.flickr.com/photos/serbianstruhpro/">https://www.flickr.com/photos/serbianstruhpro/</a>
---	----	------	---	-----------------	---

At the bottom of the web page, three buttons allow the user to download:

- This report
- The rights holders' report that the operator could generate and communicate to the rights holders in application of settlement agreements and retrocession of equitable remuneration
- The statutory report provided for by the AI act (art 53, 1, d), according to the form proposed by the Higher Council for Literary and Artistic Property (CSPLA)

## How it works

The simulator has an " advanced " button parameters " which allows the user to set the parameters they want. The following table copies those that are set by default and indicates the calculation formulas.

Parameters and values of a simulation

VARIABLE	SETTING	COMPARABLE	CALCULATION	FORMULA
User subscription	12 €/month (1)	Adobe Firefly starting at \$4.99/month Midjourney from \$10/month Dall-e included with ChatGPT Plus \$20 Getty Edify €45 for 25 generations	S	
Generated Content	60 /month	Adobe Firefly Basic Plan 100/month Midjourney basic plan 100/month	G	
Face value per generated content			0.2 €	$V = S / G$
Share of rights holders and authors				$R\% = Rr\% + Ra\%$
- Rights Holders	10%	Music (ASCAP, BMI, SACEM) 2 – 8% Stock image 55 – 85%		$Rr\%$
- Authors	5%	Printed books 5 – 15% Music (performer and composer) 15 – 30% Stock photography 15 – 45%		$Ra\%$

Remuneration of rights holders and authors per generated content		3 c€	$R = V \times R\%$
Attribution		21%	$A\% = \text{Similarity algorithm}$
		8	$N = \text{number of contents assigned}$
Training Content Reward		0.63 c€	$T = R \times A\%$
Number of authors	1,800		B
Users	500,000	Midjourney 's number of paying subscribers is estimated at between 2 and 5 million .	U
Monthly reward for rights holder and author		829 €	$S \times U \times R\% / B \times A\% \times N$
- Rights holders		276 €	$S \times U \times Rr\% / B \times A\% \times N$
- Author		553 €	$S \times U \times Ra\% / B \times A\% \times N$

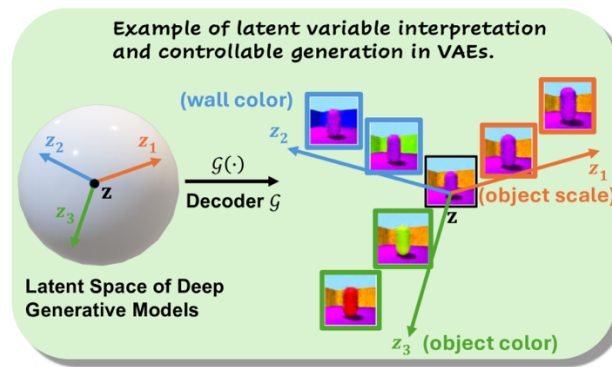
## Concept of similarity

This demonstrator uses a similarity algorithm between training and generated contents, implementing [DINOv2](#) and [FAISS](#) . Intuitively, two similar images look alike.

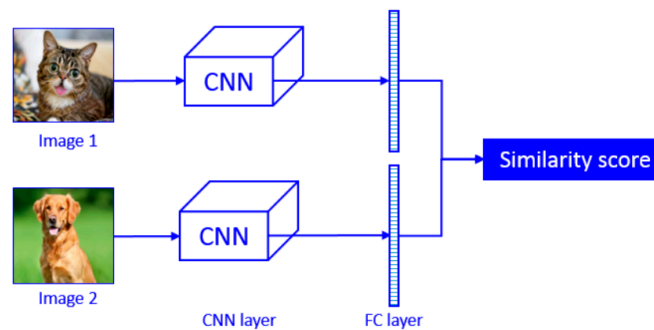
However , the "similarity" mentioned here is a scientific concept detached from any legal concept of "counterfeiting", "resemblance" or infringement of a work, defined in French law as "any reproduction, representation or dissemination, by any means whatsoever, of a work of the mind in violation of the author's rights, as defined and regulated by law" (CPI, art. L. 335-3). There is not necessarily infringement or infringement by the reproduction of an element of a work, a "contribution" in this case. In the same way, when an author creates, an element can enter into its composition, without there being infringement or infringement of a third party. The remuneration resulting from the "similarity" should therefore not paralyze the author's action in the event of infringement.

## Vector similarity

Technically, each generated or original content is analyzed to extract measurable information: cat, dog, color and size of the foreground, type, style, size, recognized objects, keywords, etc. all of these hundreds of utility indicators , around 500 in our demonstrator, form a vector in a "latent space".



The degree of attribution of a work (image 1) to a generated content (image 2) is calculated by a similarity score between the two images, processed by the two Siamese neural networks. The closer the resulting vectors are, the higher the similarity score between the images.



## Number of authors

The Open Images database was formed from Flickr's, applying filters such as non-detection by Google Image Search, a minimum number of pixels, or deduplication. Open Images also only selects CC-BY licenses, i.e. one image out of 30. As Flickr offers the "all rights reserved" license by default, the license change is only done by informed authors. This suggests a strong concentration in Open Images of the most active contributors, having published more than 5,623 photos, i.e. about 1,800 authors. This intuitive hypothesis could be verified with an exact count of the authors of the Open Images database, which we were not able to achieve given the limit in query speed.

## Identification of authors

The Open Images database identifies the authors of the images, and provides their Flickr profile page, which does not, however, include a full email or postal address.

Beyond our demonstrator and Open Images, public datasets provide identification elements, of variable precision, but which can be cross-referenced to identify each rights holder.

Cross-checking methods are best known for their fraudulent use against privacy. As early as 1997, the mandatory reporting of aggregated statistics by a Massachusetts insurer, along with electoral data, made it possible to identify Governor William Weld without any possible error and to know his health record.

Their legal use makes it possible, in most cases, to fully identify the authors: by comparing social graphs, by statistical correlation thanks to a rare combination of attributes (scientific niche, hours of activity, IP address), by reconstructing individual attributes from group averages and known attributes of other individuals, etc. The following table indicates ways to identify the authors of GPT 4 training content.



AI-generated by Adobe Firefly

Potential identification of authors of datasets used by GPT-4

	DATASET	IDENTIFICATION ELEMENTS	DIRECT CROSS-OVER	IDENTIFIER
<b>Direct identification</b>	Open Images	-	-	Name, first name, profile sheet
	Project Gutenberg			Name, first name, dates of birth and death
	arXiv	-	-	Name, First_Initial
	PubMed	-	-	Id_author
<b>Indirect identification</b>	CommonCrawl	Domain	Whois	Registrant_contact (name, organization, street, city, state, phone)
	BookCorpus	Plain_text	Smashwords	Name, first name
	WebText	Plain_text + search = "written by"	Amazon API	Name, first name
	StackOverflow	HTML	Stackoverflow.com	User_ID , network_profile
<b>In waiting for claim</b>	Wikipedia	Username	-	-
	Reddit	Username	-	-
<b>Other</b>	CommonVoice	(stats)	-	-
	Google Books Ngrams	( not applicable )	( n / A )	( n / A )
	GitHub	( not accessible )		

## 6 – Existing solutions

### Synthesis

The following table identifies some commercialized or tested methods, with regard to our demonstrator:

#### Methods for promoting author content

VALUATION	VALUATION METHOD	DEMONSTRATOR	COMMERCIAL COMPANIES	RESEARCHERS
Generated content	Subscription cost / quantity	Controversy		
Dataset	Similarity of content sets			Adobe
	Shapley			Open AI
	In- run data Shapley			Open AI
	Frequency of website access, type of data		Tollbit	
Training work	Similarity of content	Controversy	Bria (images), Prorata	
	Modeling learning		Gist.ai (texts)	Open AI
	Template customization			Adobe
	Unlearning			Adobe
	Tattoo			Adobe
	Shapley federated			Huawei
Author	Pay to train		Shutterstock , Getty , Audiosparx , Canva	
	Influence of the whole works		Sureel.ai (music)	
	Identification of rights holders	Controversy		

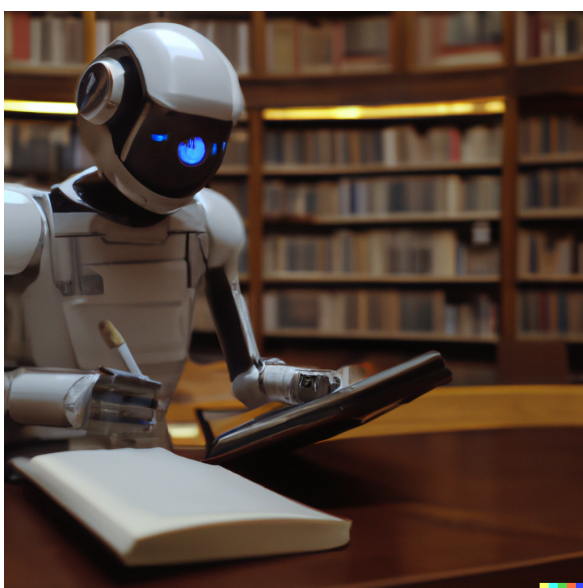


## Marketed solutions

### Companies that pay authors for training content

COMPANY	VENTURE CAPITAL	TURNOVER	METHOD	MEDIA	CONTENT PROVIDER
Adobe (Firefly)	-	\$21 billion	Pay-to-train	Picture	
Canva	-	\$2 billion	Pay-to-train	Picture	
Getty	-	\$917 million	Pay-to-train	Picture	
Shutterstock	-	\$875 million	Pay-to-train	Picture	
ProRata (Gist.ai)	\$25 million	-	Content Similarity	Text	Sky News, Time, Financial Times, The Atlantic, Axel Springer, The Guardian, dmg media, Universal Music, Fortune
Bria	\$24 million	-	Content Similarity	Picture	Shutterstock, Getty Images, Alamy , Superstock
Tollbit	\$24 million	-	Frequency of access to the website	Text	Penske Media Corporation, Time, Mumsnet, Trusted Media Brands, Candr Media Group, AdWeek
Sureel.ai	-	-	Influence of the whole works	Image, music	
Audiosparx	-	-	Pay-to-train	Music	

### Pay -to-train method



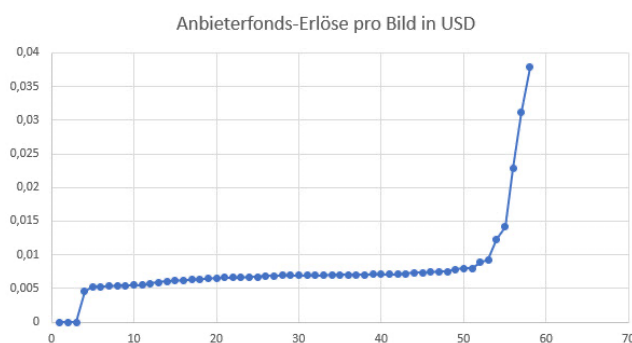
AI-generated by DALL-e

Adobe's private content banks, Canva , Getty and Shutterstock , distribute the price of the datasets they sell using the " pay to train" method. Authors are paid based on the volume of their contributions to these datasets .

pay -to-train" value can be calculated from various utility indicators. Shutterstock, for example, pays each photo \$0.0050 per year, to which is added up to \$0.0038 per photo per year depending on the amount of metadata and other indicators.



Pay -to-train value : authors' remuneration by Shutterstock . Sample N = 58 (in \$/image)



The advantage of this model is its simplicity of calculation. It has been operational for two years, and seems well accepted by rights holders.

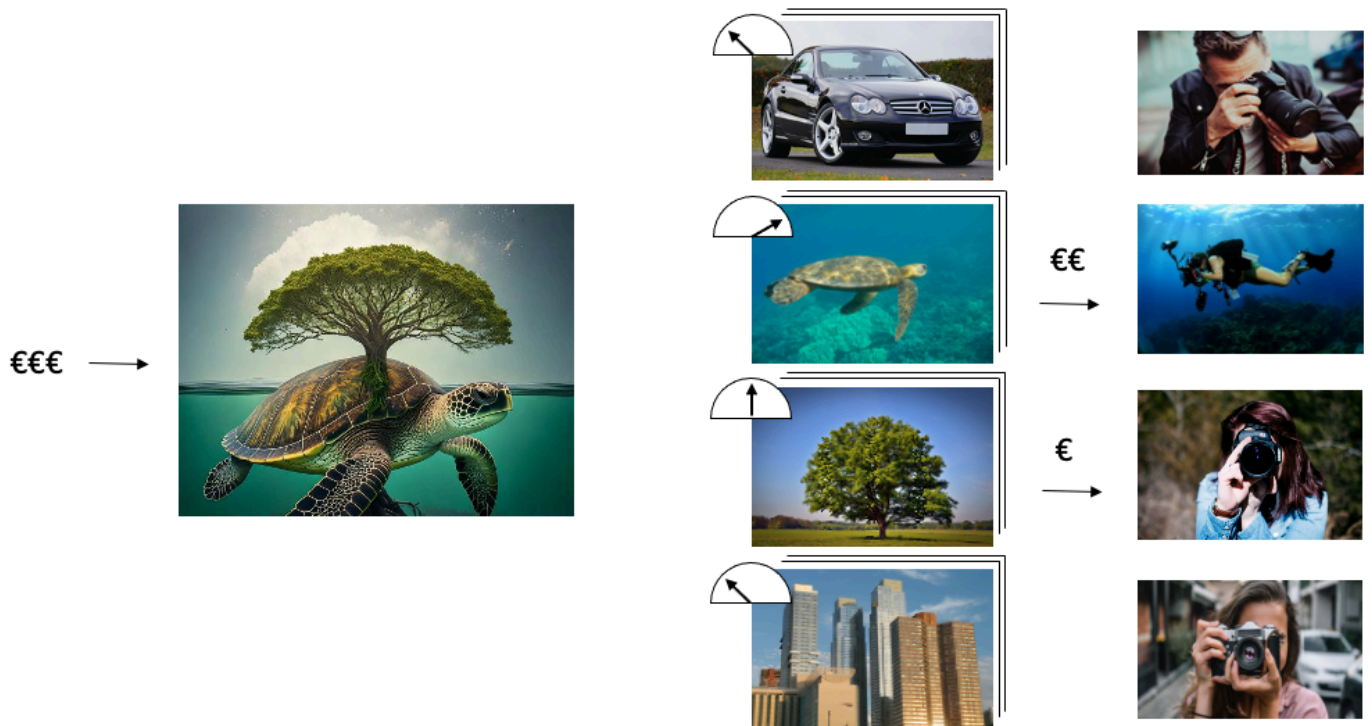
On the other hand, the level of fairness between authors presents two biases:

- (6) The quality of the content is not taken into account: an author who contributed 200 photos receives, all other things being equal, twice the remuneration of a contributor who contributed 100 photos.
- (7) usefulness for generated content is not taken into account: Van Gogh does not receive more than Picasso if users insert "in the manner of Van Gogh" more often in their prompts.

## Hybrid method Similarity + Pay -to-train

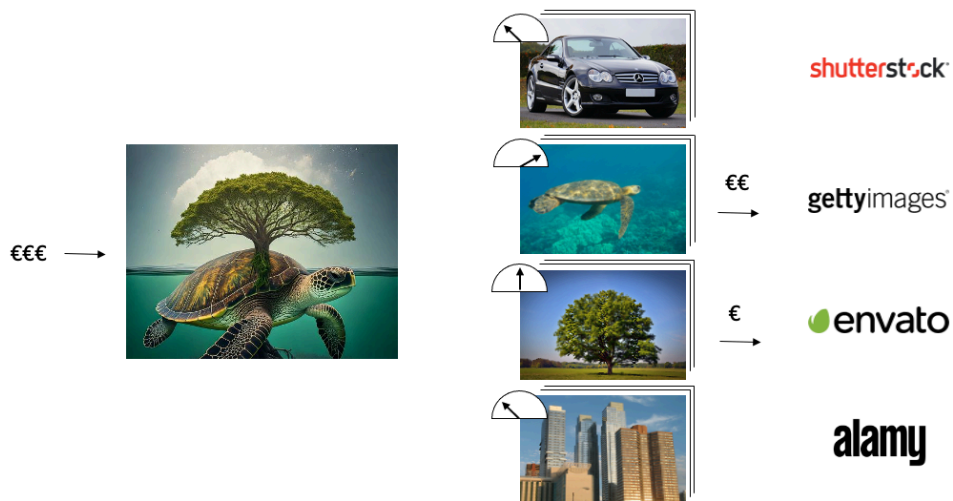
The Israeli -American startup Bria, which collaborates with large private image banks ( Shutterstock , Getty), is integrated by Big Tech (Microsoft, Amazon, Nvidia ). As a first approach, Bria uses, like our demonstrator, a similarity method, in order to remunerate the authors proportionally to the resemblance thus measured.

Measuring similarities between generated content and original photos to remunerate authors



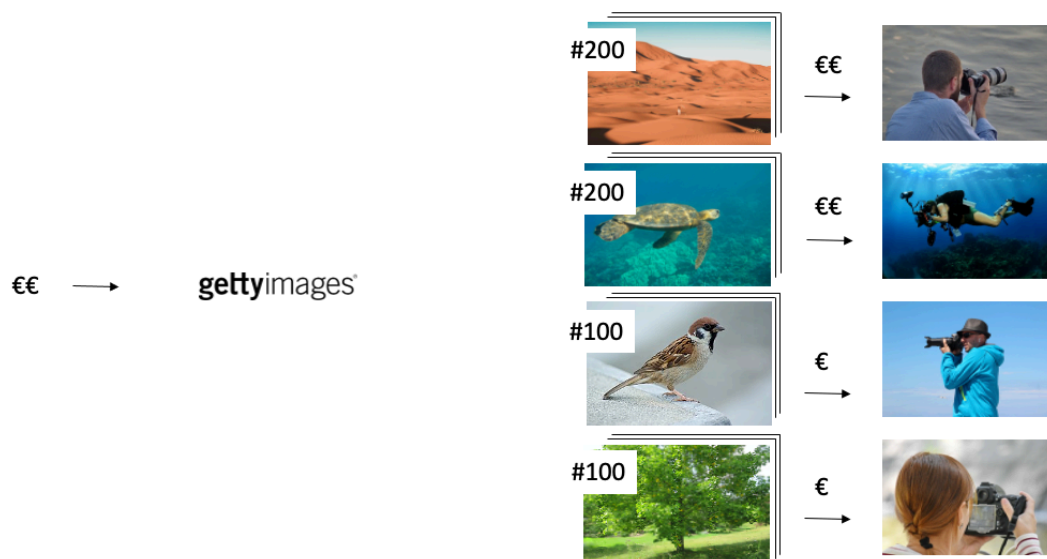
More precisely, it seems that the sum of the similarity scores is calculated by image bank ( Shutterstock , Getty, Alamy ...), which causes an initial remuneration of the banks proportional to these sums.

Measuring similarities between generated content and original photos to remunerate image banks



These banks then distribute this first remuneration between the authors, according to the individual number of their works contributing to the datasets sold, according to the “ pay to train” method .

Number of photos to pay authors



Further refinements are made, with images with quality metadata receiving a bonus.

Bria's method is therefore done in two stages:

- First, thanks to a similarity algorithm between training images and generated images
- Then, thanks to an algorithm of quantitative contribution of the authors to the datasets sold (“pay - to-train”)

This hybrid solution is:

- economically compatible with the business model of generative AI
- accepted for two years by image banks and authors
- currently being extended to other types of content: 3D, music, voice, animated image, video

It's possible that startup ProRata , in particular, has adopted this hybrid method and is offering it to media outlets that already have deals with AI operators. It involves “taking the output of generative AI, whether it's text, image, music, or film, and breaking it down into its components, figuring out where they came from, and then assigning a percentage to each copyright holder, and paying them accordingly .” Its parent company's public patents don't provide any more details.

This hybrid method still has some flaws:

- It is not proven that this solution will adapt to all types of content, for example games
- There is no guarantee that the similarity is not coincidental. In other words, an author could be paid even though his works were not useful for the generated content, and vice versa.
- The contribution of an original image is measured roughly. In the example above, the desert photographer is paid the same as the sea turtle photographer because their number of contributions to the Getty Images dataset is the same, while the desert photos are of zero usefulness to the generated content.

It should be noted that:

- even if the distribution of compensation for a particular generated content is approximate, its fairness probably improves with the quantity of generated content.

- individual identification of authors by Bria is not necessary, but only identification of training datasets
- The distribution algorithm itself is not public. It is possible that the authors will demand more transparency in the future, but it is also likely that a dose of secrecy will be necessary to avoid opportunistic biases.
- We also do not know the average rate of remuneration of authors in relation to the turnover of the content generated, which, in our opinion, could be published.

## Dataset Similarity Method

Startup Sureel provides musicians with a tool that measures the level of influence their work has, measured in percent, on an AI-generated track.

The algorithm, which he describes in his '949 patent, provides for an author's compensation based on the influence of his or her entire body of work, measured for example by the difference between the weights of the base model and those of its fine-tuning by that ensemble. This variation aims to dilute the similarity bias by comparing a larger number of works.

## Methods currently being tested

Beyond these few variations, researchers are competing in ingenuity to test new methods:

Search for fair compensation models

BUSINESS	METHOD	RESEARCH TEAM	DATE	MOTIVATION
OPEN AI Princeton University Columbia University Harvard University University of Pennsylvania University of California Berkeley	Shapley	<b>An Economic Solution to Copyright Challenges of Generative AI</b> Jiachen T. Wang, Zhun Deng, Hiroaki Chiba-Okabe, Boaz Barak, Weijie J. Su	April 14, 2024 September 9, 2024 (v4)	Shapley values provide a principled approach to distributing gains based on utility.  The utility measures the extent to which data sources are responsible for generating content. The value is low if the counterfactual model is unlikely to generate the same content as the deployed model, and vice versa.  Utility can be interpreted as the total compensation that all members collectively deserve for providing their data to train the generative AI model.

BUSINESS	METHOD	RESEARCH TEAM	DATE	MOTIVATION
Virginia Tech Massachusetts Institute of Technology	In-run data Shapley	<b>Data Shapley in One Training Run</b>  Jiachen T. Wang, Prateek Mittal, Dawn Song, and Ruoxi Jia	June 29, 2024	Boaz Barak, a Harvard professor of safety ("superalignment ") who also works at OpenAI , doesn't lose sight of Shapley's ideal method, even if it means compromising on accuracy.  His team of researchers demonstrates in real life how to calculate the Shapley value of datasets like Wikipedia or Github for a general-purpose AI model trained from a GPT-like dataset.  According to this team, aligning remuneration with the quantifiable contributions of each dataset guarantees the best fairness in the distribution of royalties. . She indicates her preference for the less expensive In- run data Shapley variant.
	Training modeling	<b>TRAK: Attributing Model Behavior at Scale</b>  Sung Min Park, Kristian Georgiev , Andrew Ilyas, Guillaume Leclerc, Aleksander Madry MIT	April 3 , 2023	Aleksander Madry , a professor at MIT, is the lead for AI reasoning at OpenAI . Trak's solution relies on an approximate representation of the AI model to describe how it evolves during training and connects predictions to the training data that influenced them.  This solution has proven useful with image classifiers trained on ImageNet , vision-language models (CLIP), and linguistic models (BERT and mT5).
ADOBE  Carnegie Mellon University  UC Berkeley	Dataset similarities	<b>Evaluating Data Attribution for Text- to-Image Models</b>  Sheng-Yu Wang, Alexei Efros , Jun-Yan Zhu, Richard Zhan	August 8, 2023	Since the similarity method risks revealing fortuitous resemblances, this bias can be diluted by extending the method to two sets of images, training and generated. The Adobe researchers thus calculate the contribution of the first set, the dataset , to a set of synthetic contents
	Watermarking	<b>ProMark : Proactive Diffusion Watermarking for Causal Attribution</b>  Vishal Asnani , John Colomosse et al.	March 14, 2024	Watermarking demonstrates causality between training content and synthetic content. This allows for confident legal attribution or payments to creators.
	Unlearning	<b>Data Attribution for Text-to-Image Models by Unlearning Synthesized Images</b>  Sheng-Yu Wang, Aaron Hertzmann , Alexei A. Efros , Jun-Yan Zhu, Richard Zhang	June 2024	This method unlearns the synthesized image, by training the forgotten images by proxy, identifying those with significant loss deviations after the unlearning process, and labeling them as influential.
HUAWEI  University of British Columbia  Simon Fraser University  KTH Royal Institute of Technology	Federated Shapley	<b>Improving Fairness for Data Valuation in Horizontal Federated Learning</b>  Zhenan Fan, Huang Fang, Zirui Zhou, Jian Pei, Michael P. Friedlander, Changxin Liu, Yong Zhang	May 23, 2022	The motivation of data owners depends in part on the fairness of collaboration and reward in federated learning.

## Variants and developments

In all the methods mentioned:

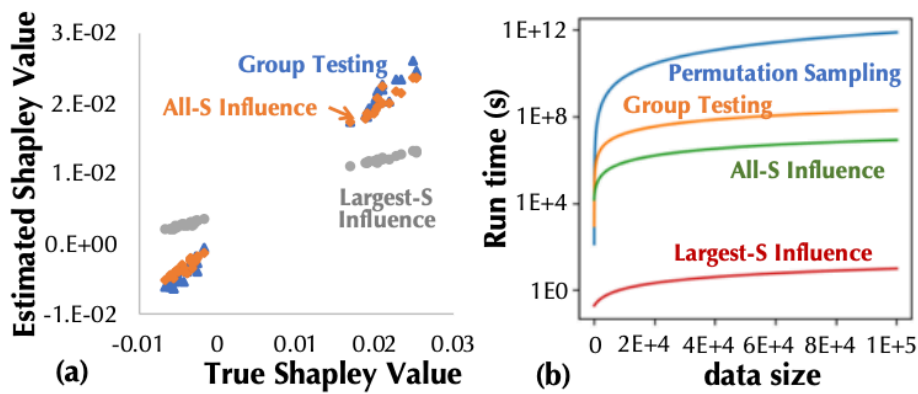
- The remuneration parameters are adjustable based on arbitrations on the utility indicators: size, date, similarity, number of metadata, independent commercial success, etc.

Indicators of utility

RESEARCHER, COMPANY	MEDIA	UTILITY INDICATORS
Piao , Chen et al.	Picture	Vectors
Wang, Deng et al.	Encyclopedia, Academic publication, Image, Code	Counterfactual model (S) Contents generated by the x( gen ) model and the counterfactual model
Wang, Mittal et al.	Encyclopedia, Academic Publication, Image, Code, Email, Forum	Hold validation data point set Z( val )
Deng, Zhang et al.	Music	Average speed of events Average height of all events Sum of time differences of all events
Shutterstock + Bria	Image, vector image	Similarity score Revenue by dataset Data volume per dataset and per author Metadata volume Role played in the development of the original models Role played through royalty payments linked to future generative licensing activity
Adobe Stock + Bria	Image, vector image	Similarity score Image admissibility Image date Number of downloads

- The algorithm is tunable between accuracy and execution speed. The Shapley value, in particular, can be calculated
  - with an arbitrary number of generated contents
  - with arbitrary granularity (by dataset , sub- dataset , down to individual data)
  - with similar functions

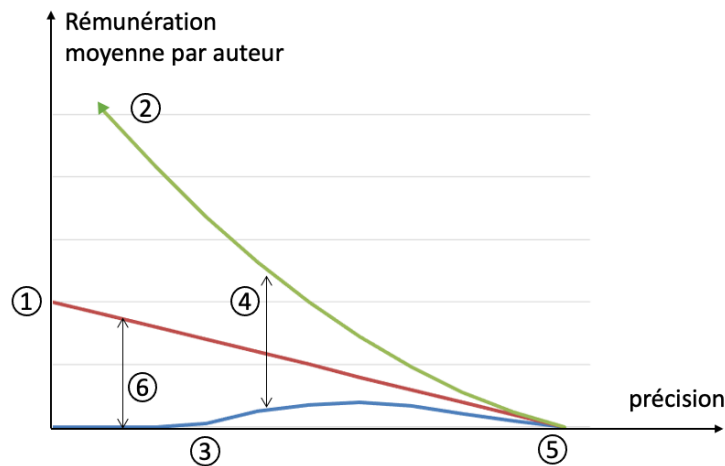
Three functions approximating the Shapley value



At this stage of the state of the art, the field of investigation is sufficiently broad to hope to find remuneration solutions, by use, which optimize the relationship between equity, cost and precision.

Impact of the cost of precision on remuneration.

- (7) At zero precision, the remuneration per author is equal to the total collection divided by the number of authors.
- (8) At zero precision, the maximum potential remuneration of an author is equal to the total collection
- (9) At zero precision, the minimum potential remuneration of an author is equal to zero
- (10) Uncertainty gap
- (11) When the cost of precision exceeds total collection, the compensation and uncertainty gap are zero
- (12) At constant collection, the average remuneration per author decreases with precision



## 7 – Acceptability



**#FairlyShare contre le péché originel d’Internet** – septembre 2014, par Vincent Lorphelin, Gilles Babinet et al.

Ce qu’ils dénoncent ? Les plate-formes qui reposent sur l’industrialisation du travail gratuit, en dehors des règles de l’appel d’offre, du concours ou du bénévolat. Ils emboîtent ainsi le pas à [...] la levée de boucliers qui a suivi le rachat du journal de crowdsourcing Huffington Post par AOL pour 315 millions de dollars. [Lire](#)

pay -to-train” method is currently the most widespread for remunerating authors whose works are used to train general-purpose AI models. Despite partial transparency and fairness, it has not yet given rise to any lawsuits or petitions, as we saw, for example, during the awareness of “free work” or the sale of the Huffington Post (see box).

Media banks using “ pay -to-train” and AI models using their datasets

DATE	MEDIA	DATASET PROVIDER	AI MODEL PROVIDER
10/25/22	Image, photo	Shutterstock	OpenAI
01/12/23	Image, photo	Shutterstock	Meta , Google , Amazon , Apple
03/12/23	Image, photo	Shutterstock	LG ( Exaone )
03/21/23	Image, photo	Getty Images	Nvidia
05/29/23	Avatar	Ascendant Art	
07/11/23	Image, photo	Shutterstock	OpenAI
08/08/23	Image, photo	Shutterstock	Nvidia (Picasso)
09/07/23	Image, photo	Getty Images Alamy Bria	

DATE	MEDIA	DATASET PROVIDER	AI MODEL PROVIDER
09/13/23	Music	AudioSparx	Stability AI
09/13/23	Image, photo	Adobe Stock (Firefly)	
09/26/23	Image, photo	Getty Images (Getty Gen AI)	
10/04/23	Image, photo	Canva (Magic Studio)	Open AI (Dall-e) Google (Image)
02/21/24	Image, photo	Superstock k Envato	
04/11/24	Image, photo	EyeEm Freepick Photobucket	
06/27/24	Music, Voice	Universal Music* Sony Entertainment Warner Records	Google ( Youtube )
07/30/24	3D	Shutterstock	Nvidia (Edify)

\*experimentation



Dataset purchase contracts from media banks have been accompanied by speeches that promote the “respectful, responsible, secure and fair” remuneration of authors.

### Speech on the fairness of pay -to-train

PLATFORM	STOCK AND DATASET	DECLARATIONS
Nvidia	Getty, Shutterstock, Adobe	Nvidia : “We’ve built a platform that allows our partners to train on licensed data, for example from Getty, Shutterstock or Adobe,” Huang said. “They respect the content owners. The training data comes from that source and whatever economic benefits come from that can go back to the creators.”
	Adobe	Adobe: We are developing generative AI responsibly, with creators at the center . Adobe: Create with confidence knowing your output is of the highest quality, built with templates designed to be safe for commercial use.
	Getty	Getty: Generative AI by iStock is an affordable and commercially safe generative AI tool [...] without the fear that something legally protected has slipped into the dataset and could end up in their work
Meta	Shutterstock	Shutterstock : Expertise in building a large-scale ecosystem that rewards and connects contributors to creators.
LG Exaone	Shutterstock	Shutterstock : Ownership of AI-generated content cannot be attributed to an individual and must instead compensate the many artists who participated in the creation of each new piece of content. Our customers can safely and legally license AI images for their own purposes, without worry. We also compensate the contributors whose work trains our models, so you can use our AI with a clear conscience, too. Our AI image generator remains the safe, easy-to-use, and responsibly designed feature you know and love.
		LG: The AI model is trained with millions of high-resolution images and metadata from Shutterstock and will convert text prompts into images. Contributors whose works were used to train the model will be compensated through Shutterstock ’s Contributor Fund and will also be compensated each time new generative content using their IP is created and licensed by customers.
Open AI	Shutterstock	OpenAI : The data we licensed from Shutterstock was essential to the formation of DALL-E. Ensuring that the creator economy continues to thrive is a major priority for OpenAI . Writers, artists, composers, and other creators have contributed immeasurably to societies throughout the history of civilization [...] OpenAI does not want to replace creators .
	Canva	Canva : Magic Media ’s text-to-image converter , OpenAI ’s DALL·E , and Google Cloud’s Imagen [...] ensure their AI models are fair and ethical. “We share the value we build in Canva with the creators who have been with us for all these years. With the new Creator Fund , they have access to a royalty pool that matches their AI creations.”
Google Cloud	Canva Shutterstock	Google Cloud: Shutterstock also became a leading innovator by being the first to launch an ethically sourced AI image generator, now enhanced with Imagen on Vertex AI.
Google Youtube	UMG	Google subsidiary YouTube is trying to strike deals with music majors for the use of AI. It wants to ensure that " music rights holders are compensated for their contributions to training data. UMG and YouTube are in talks to license artists' vocals and melodies to train AI models."
Amazon SageMaker	Alamy	Alamy : At the heart of the Fair Diffusion programme’s ethos is a commitment to fairly remunerating photographers, artists and creators.
	Getty	Getty: We adhere to the highest ethical standards and respect the intellectual property and privacy rights of others.
	Envato	

PLATFORM	STOCK AND DATASET	DECLARATIONS
----------	-------------------------	--------------

Sony

Sony's president said: "New products and businesses created from music must be developed with the owner's consent and appropriate compensation and credit. It is essential to understand why AI models are being trained, what products will be developed as a result, and what business model will monetize the use of the artist's work."

"If a generative AI model is trained on music for the purpose of creating new musical works that compete in the music marketplace, then the training is not fair use. In this case, the training cannot be done without the consent, credit, and compensation of artists and rights holders."

## 8 – Patents

In preparing this report, we identified the following patents:

Patents relating to the remuneration of generative AI training works

TITLE	REF	COUNTRY OF THE DEPOSITOR	DELIVERED	PIONEER
Systems and methods of creative work collaborative systems	US9100444	US	✓	✓
Stakeholder equity valuation in collaborative projects	US10776889	US	✓	✓
Attributing generated visual content to training examples	US12033372	US	✓	
Model-based attribution for content generated by an artificial intelligence	US12013891	US	✓	
Generating author vectors	US10599770	US	✓	
Input-based attribution for content generated by an artificial intelligence	US2024419949	US		
Output-based attribution for content generated by an artificial intelligence	US2024419720	US		
Fee Distribution System and Method	US2023401550	US		
Generation, usage and attribution of content	WO2024166091	US		
Facilitating generation and usage of visual content	WO2023084512	US		
Federal learning-based online client selection method and system, medium and equipment	CN118446287	China	✓	
Image index generation based on similarities of image features	CN103207879	China	✓	
Method and system for evaluating contribution degree of image local features to overall features	CN111652239	China	✓	
Method and system for quantifying client contributions in federated learning	WO2023036184	China		
Federal learning mobile device distribution data processing method based on Shapley value	CN114912626	China		
Federal learning excitation method based on cooperation Sharpley value	CN115965092	China		
Methods and systems for federated learning with local predictors	WO2024031564	China		
Method and collaboration system	US 9723059	EU	✓	✓
Method for an explainable autoencoder and an explainable generative adversarial network	US11948083	EU	✓	
Architecture for explainable reinforcement learning	US11455576	EU	✓	
Multistage contribution traceability in collective creation environment	US2015363743	EU		
Method and device for rewarding collaborative creation	EP2958068	EU		

This list shows that, when it comes to remunerating rights holders, Europeans are not outdistanced by China and the United States, which contrasts with their outsider position for artificial intelligence in general and generative AI in particular.

### The EU's position in complex (digital and green) technologies

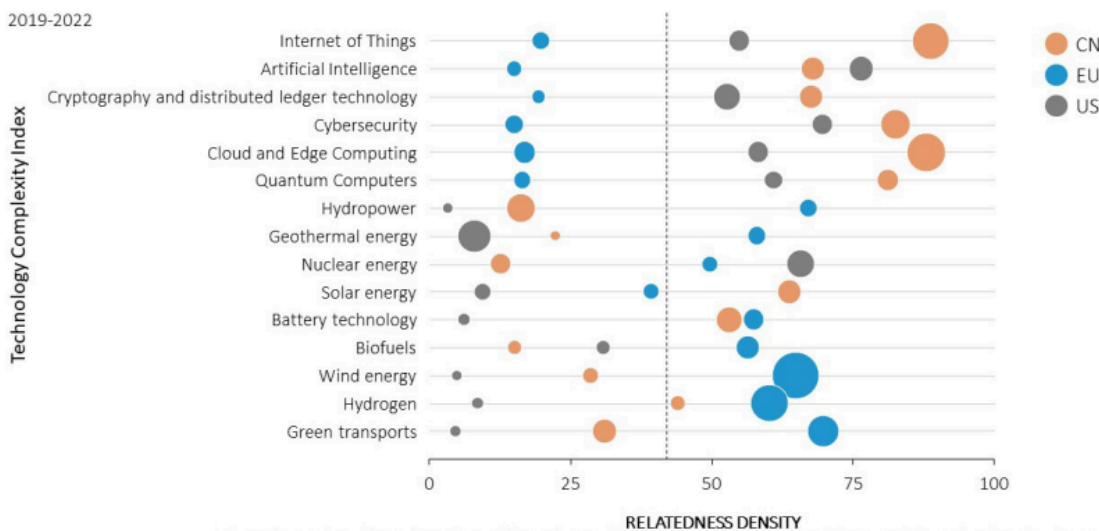


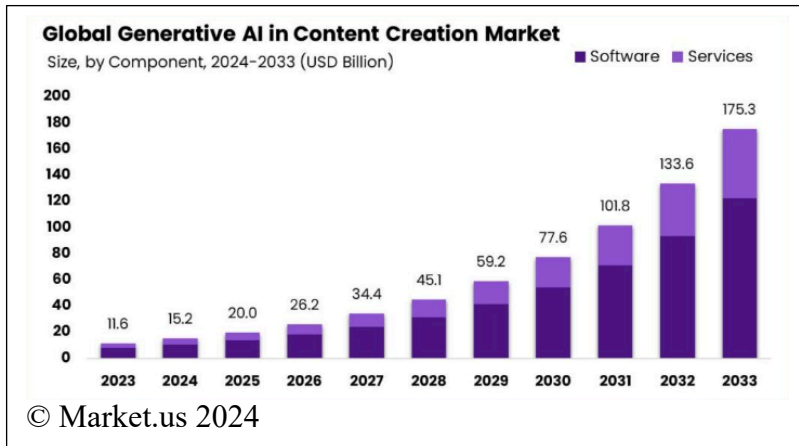
Table 7 Top patent owners in GenAI modes (companies), 2014–2023

	Images/ video	Text	Speech/ voice/ music	3D image models	Molecules/ genes/ proteins	Software/code	Other modes
Tencent Holdings (China)	607	565	551	102	57	41	464
Ping An Insurance (China)	262	600	599	26	33	24	223
Baidu (China)	395	465	441	81	26	18	166
IBM (US)	101	274	168	13	37	42	132
Alibaba Group (China)	142	213	144	36	2	10	133
Samsung Electr. (Republic of Korea)	173	140	226	34	10	5	57
Alphabet/Google (US)	138	107	200	25	15	15	72
ByteDance (China)	173	82	112	21	6	7	90
Microsoft (US)	78	194	151	18	6	22	42
BBK Electronics (China)	125	58	117	12	0	4	116
Netease (China)	78	76	80	37	0	11	102
NTT (Japan)	37	100	169	11	0	3	62
Huawei (China)	104	84	72	24	1	4	102
China Mobile (China)	79	67	83	10	0	7	89
State Grid (China)	75	54	31	6	1	2	144
Adobe (US)	190	79	34	30	0	9	20
Sony Group (Japan)	87	35	105	42	0	1	42
Siemens (Germany)	124	25	8	10	3	4	59
Ant Group (China)	45	82	29	3	0	6	59
Industrial and Commercial Bank of China (China)	40	62	49	5	0	16	48

Source: WIPO, based on patent data from EconSight/IFI Claims, April 2024.

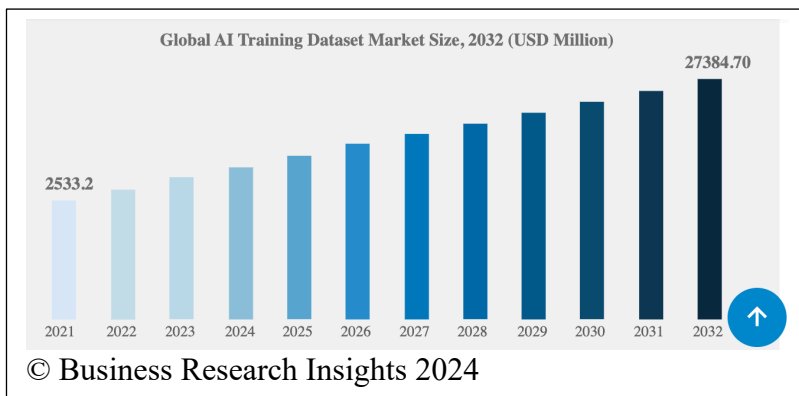
# Perspectives

## 9 – Market



For the creativity segment alone, the global generative AI market size is estimated to reach USD 11.6 billion in 2023 and will exceed USD 100 billion by 2031.

To give an idea of the scale, a fair remuneration rate of 15% would represent more than the current collection of OGCs (Collective Management Organizations) for artistic rights (music, audiovisual, live performance, literature and visual arts).



Companies specializing in AI training datasets organize the collection, annotation, conservation of data and their adaptation to the needs of learning algorithms.

The global AI training dataset market was \$2.5 billion in 2021 and is expected to reach \$27 billion by 2032.

### Compensation models for generative AI data authors

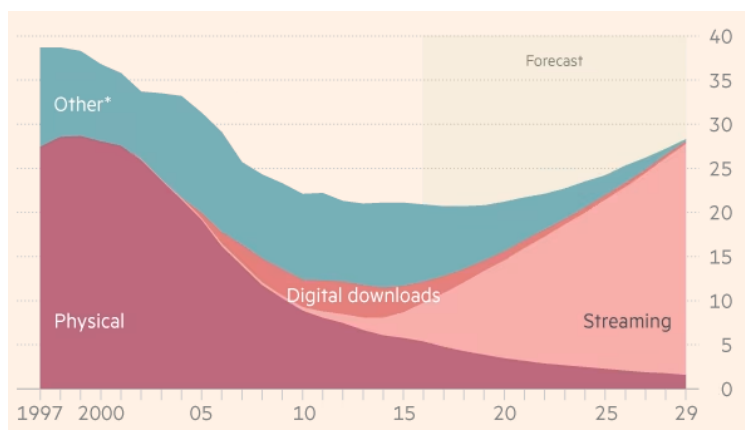
BUSINESS MODEL	COMPANIES	AUTHOR COMPENSATION
Scraping (1)	OpenAI , Midjourney , Stability, MistralAI	No
Uber of data (2)	Mturk , ClickWorker , Gigwalk , Premise	Payment on order
Content banks (3 )	Shutterstock, Getty, Canva, Adobe	Pay-to-train
Media groups (4)	Axel Springer, Financial Times, Stack Overflow, Reddit, Le Monde	No
Companies of authors (5)	the Authors' Guild , Created by Human, Human Native AI, Fairly Trained, Dataset Providers Alliance	Fair pay

- (3) Scraping : DALL·E, for example, was first trained with hundreds of millions of captioned images from the Internet. This practice, which is the majority, is the source of many lawsuits.
- (4) Uber of data: AI model makers have begun to hedge risks and secure data supply chains, both through a thriving industry of data brokers and through deals with content owners that have sprung up to meet demand.  
An industry of specialized AI data companies is emerging, creating networks of short-term contract workers to produce custom content on demand. Prices range from \$1 to \$2 per image, \$2 to \$4 per short video, and \$100 to \$300 per hour for longer films. The market rate for text is \$0.001 per word. Owners of photos, podcasts, and medical data are paid about 20 to 30 percent of the total transaction value.
- (8) Content banks: Shutterstock , Getty, Canva or Adobe have large media bases supplied by authors, paid according to the datasets sold. Shutterstock , leader in images, has sold licenses to AI providers for more than \$100 million, or 12% of its total turnover, and is expected to increase by 40% in 2024. These image, video and music licenses will form an addressable market of \$10 billion in 2030, growing annually by 22% over the period.
- (9) Media groups: Database owners license datasets , without additional compensation <sup>50</sup>authors
- (10) Author societies: authors want to be able to define their licensing options for generative AI. Ideas are flourishing like that of *the Authors' Guild with the Created by Human* platform , like *Human Native AI* , *Fairly Trained* , *Dataset Providers Alliance* or the *AI Commission* in France. At this stage, the economic models are not defined and the development prospects are not quantified.

## Licensing proposals for works for generative AI

INITIATIVE	MAIN PROPOSALS
The Authors' Guild	<ul style="list-style-type: none"> <li>- Establishment of agreements for the use of works</li> <li>- Payment of royalties for the use of protected works</li> <li>- Transparency on the data used</li> </ul>
Created by Human	<ul style="list-style-type: none"> <li>Human-Created ” label for works</li> <li>- Platform allowing authors to authorize and control licenses</li> </ul>
Human Native AI	<ul style="list-style-type: none"> <li>- Marketplace for generative AI training datasets</li> <li>- Tracking permissions for each content used</li> </ul>
Fairly Trained	<ul style="list-style-type: none"> <li>- Label “ Fairly Trained » for generative AI</li> <li>- Standardized contracts for the use of works in AI</li> </ul>
Dataset Providers Alliance	<ul style="list-style-type: none"> <li>- Fair remuneration</li> <li>- Standardized contracts ( Pay -to-train, royalties on turnover, subscription, license by sector)</li> <li>- Transparency on the use of training data</li> <li>- Individualized authorizations by author</li> </ul>
AI Commission	<ul style="list-style-type: none"> <li>- Public platform</li> <li>- Intermediation of the National Library of France and the National Audiovisual Institute</li> </ul>

Drawing a parallel with the music market, we are currently in the "Napster -IA", waiting for solutions equivalent to Spotify and Apple Music. Their business models have cleaned up the market since 2012, where streaming and piracy were previously confused.



© Financial Times, Goldman Sachs

voluntary exploration by stakeholders.

Despite their embryonic nature, the multiplication of these initiatives opens up a new perspective of "liquid property".

Similarly, we believe that the potential market for the valorization of works by generative AI is greater than that of OGCs, and represents an opportunity for rights holders. Symmetrically, generative AI operators could access quality content bases.

This potential, which for the moment benefits alternative segments of the market, will be released by the proposal of new economic models of the "authors' societies" type and their



## 10 – Towards liquid property

### Interview de Vincent Lorphelin

Le Soir (Bruxelles), Janvier 2024



"Sam Altman and Elon Musk say that generative AI will make work obsolete , so we must prepare for universal income. This coagulates with the thesis of the end of work relayed by Jeremy Rifkin a few years ago . It also joins that of the singularity which advances that in a few years , general AI go surpass humans in all areas. The proof is that we have always said that AI will never enter the professions creative (she will never be an author, designer...). Except that since last year, we are there. The doors of the bastion are broken down . [...]

Conversely, the European DNA is the social market economy , it is the economy that makes society . Work makes sense. Historically and from a strategic point of view , the collective management of rights is twice as developed in Europe as in the United States . Therefore, if we come back to this

fundamental question of the interest general , to counterbalance the singularity thesis , the interest general European is to promote the model decentralizes .

Let's imagine that this decentralized platform model becomes systemic , becomes a system , in the same way that centralized platforms like Gafam do today . In this case, the matter manipulated , these are intellectual rights, not commodities. They are non-market economic value . The economist Karl Polanyi had argued that the market economy had commodified labor, whereas labor cannot be equated with a commodity.

Centralized platforms , on the other hand, will try to give a market value to intellectual rights as much as possible. In a decentralized logic, rights will be expressed in percents, not in currency. Economic values manipulated are no longer expressed in euros or dollars. It is an economy that is based not on money, but on property . To draw a parallel with what Emile Zola called "liquid money" (with the transition from an economy where wealth no longer came from land ownership but from the circulation of money), we propose the notion of " liquid property " , where the working capital of the economy is the circulation of intellectual property rights .

Today, we have financialized everything so much that a company presents accounts as if it were for sale every day, because it measures everything in money. However, there are intangible , off-balance sheet assets that express values that are not expressed in money. The wealth of the company exceeds what money expresses. As soon as we manipulate liquid properties , we are led to find other indicators. Gradually, we are evolving towards a society that makes wealth of another order solvent." (read more [here](#) ).

Just as Uber's platform model inspired the concept of uberization , we call " sacémisation " the phenomenon of the diffusion of the economic model of copyright societies, which will enhance human capital through liquid property.

## 11 – Promote human intelligence

Brad Smith, Microsoft's president, is committing a moral fault against society. Speaking of generative AI like ChatGPT, he says, "We all have a right under copyright law to read and learn. Now we're asking ourselves whether we can allow machines to learn in the same way." manner. I think there is an imperative societal to make this possible."

This intervention may seem harmless. After all, if machines absorb the Internet to create articles and illustrations at a ridiculous price, why deprive themselves of it? Aren't the New York Times and other rights

holders, who oppose this use of their content, making a bad case against Microsoft? If machines do not plagiarize, where is the harm? It is in the unacceptable purpose of this speech.



### An advantageous legal vagueness

Gafam platforms harvest a mass of small contributions and promote free work. They change dimension again with AI, whose voracity requires the conquest of new territories, those of the authors.

As an exception, the legislator had defined "fair use" to authorize Google to publish excerpts of private content. Since the legal outcome is still uncertain, former lawyer Brad Smith knows that public opinion will play a big role. This is why he distils the adulterated mental image of the "right of machines to learn." His "imperative" societal "is only a vulgar contestation of the

legitimacy of authors to defend their territories.

He distracts by debating falsely ethical issues. Is human intelligence doomed to be downgraded? Can we let AI make decisions? He may well adopt the posture of the wise man in front of an Elon Musk less polite, they are both augurs of the same illusion: that of surpassing human intelligence.

### AI: an illusion of superiority ?

The best novel will never replace the experience of life. The best "visio" will never replace a real encounter. However perfected it may be, the model is not reality. Each of us has an intelligence of the world that will always surpass that of machines, which have no intelligence other than that of their models.

AIs illuminate our intelligence. They allow the environmental police to detect polluters, the teacher to understand the difficulties of a thousand students, or the laboratory to anticipate the effects of a drug.

Spreading the illusion of the imminent superiority of machines is irresponsible. It is to deny strictly human work, which draws its inspiration precisely between the world and its models, when the artist reveals

something unsaid, when the driver compensates for poorly responding levers , when the doctor doubts a diagnosis that is too obvious , or when the judge forms an intimate conviction.

### The defense of authors

Feeding the illusion of a competition between intelligences and attributing human rights to machines is a fault against society . On the contrary, we must affirm the preeminence of human intelligence and guarantee the integrity of its territories.

Brad Smith's illusion would not be so serious if it were not the work of the world leader in AI. Won over by this illusion, the Interministerial Commission on AI claims that machines will surpass humans. It defends , fortunately, the fair remuneration of human creation , but without realizing that it is the appreciation even of this accuracy which is manipulated .

Europe has already decided with the AI ACT to protect copyright. It must now establish the land register of the territories exploited by AI and defend their rent. ( [Read](#) more in Les Echos of May 6, 2024)

## 12 – A sacemizing solution for Europe

### Entretien avec Jean-Paul Betbèze

Economiste, Membre du Comité scientifique de la  
Fondation Robert Schuman



### What should be the key points of “**sacemization**”?

sacemizing ” solution for Europe should:

- Be based on the intellectual property of the authors, like the OGCs. It is necessary to maintain and guarantee the intellectual property of the authors, which could otherwise be dissolved or disappear in the current revolution of communication. Communication is becoming globalized, it is the basis of the major change that we are experiencing, with the multiplication of messages in all directions. The contribution of each must be recognized, unless we dry up the current global movement by excessive concentration for the benefit of certain authors, the others finding no advantage in their work and their ideas, no recognition, no remuneration. Because of the gaps in the availability of capital, to the advantage of the United States for a large part, the danger is a monopolization of the apparent sources of creation.

- Collect royalties on the exploitation of works by AI. “ Sacemization ” not only allows to trace, but also to identify and estimate the sources, even when they seem minimal. This process allows to preserve all the origins and therefore all the recognitions of creations and improvements: nothing is lost, everything is kept.
- Recognize creativity and distinguish it from amateur content. In the proliferation of works, it is important to be able to highlight those that are original, to signal and promote them.
- Implement individual incentive mechanisms, so that the price of creativity is competitive on the market and compatible with the cost of production. The risk of the multiplication of works is trivialization, especially since their physical production cost is obviously low. Invoicing linked to the cost of production alone runs the risk of being disincentive . The great advantage of “ sacemization ” is to move away from a logic of price linked to the cost of production, to move towards another logic, that of the markets, which is based on the importance of success, the repetition of the message.

### **What is the advantage of a market driven by “ sacemization ”?**

In this market logic, it is measured demand, from which anticipated demand arises, which creates new prices, and therefore margins. In this context, margins can only increase, and therefore marginal profits, and therefore valuations. This difference between low costs, detected by search engines, and valuations which will thus have an objective basis, updated by Sacem -type tools , is the basis of the proposed logic. It will make it possible to highlight the contributions of each, particularly in a European context.

### **And for creative activities?**

Let us not forget that monopolistic competition is always at work: it pushes towards reducing the number of competitors, until only one emerges: it is the monopoly. There more than anywhere else, the winner takes all the stakes. “ Sacémisation ” must oppose this by allowing for better distributed remuneration before a single monopolist appears. This is obviously favorable to research, innovation, and creation, since we know well that the monopoly situation turns against the monopolist itself, in a market situation that dries up in the long term, because its dynamics lead to a drop in its profits. “Sacémisation ” thus respects creations and slows down the merciless and, ultimately, destructive selection of the monopoly itself.

In the market of ideas, speed is deadly, the memory of contributions is decisive, in connection with their fair remuneration. It is possible today and, we understand, more necessary than ever. It is about saving creation thanks to the progress that technology offers.



Michelle Bergadaà, Professeure émérite à l'université de Genève.

Co-autrice de *Réinventer l'intégrité académique à l'ère de l'intelligence artificielle* (à paraître, EMS, février 2025)



Generation Z) allows the production of not only false information, but also false authors, who are just as credible. Second, the eternal present of communities of yesteryear maintains an alternative world where the earth is flat and bleach eliminates Covid . Third, linear time is driven by Elon 's materialistic technology. Musk . AI is a fraud accelerator that thrives in the interstices of these temporalities.

### How to act?

Pubpeer and Retraction Watch Database seek to catch individual offenders, one by one. But it takes three years to complete the investigation and get the newspapers to retract.

This phenomenon must be tackled systematically with integrity radars, an injunction for fraudsters to return the money received, and exclusion AI databases. The future of our society is the young people we train. Let's help them by taking a firm stance and by paying for their original creations through generative AI. This would motivate many young, honest researchers to create knowledge."

### You say that the AI Act does not only concern authors, but also researchers, why?

MB: "We must consider applying the AI Act to remunerate the authors of "authentic research". France pays its researchers poorly, this would slow down their flight abroad or to the private sector. It costs society much more to crush its young researchers under bureaucracy than to pay them well. Many become demotivated and do overtime teaching instead of doing research. So if the State cannot remunerate them properly, the system for producing articles, books, reports must do so. This must be the cornerstone of the European strategy for generative AI. With the added bonus of an automated fraud detection system.

### Hasn't fraud always existed?

The opioid crisis in the US [Fentanyl and related drugs, prescribed by prescription, alone were responsible for approximately 71,000 drug overdose deaths in 2021] came about through a single, false, misquoted, and reprinted article. This was long before AI. With AI, the spread of fraudulent articles is primarily via new articles. This is a new model for the swarming of fakery.

Today, three temporalities coexist. First, instantaneous and emotional time (of Trump and Generation Z) allows the production of not only false information, but also false authors, who are just as credible. Second, the eternal present of communities of yesteryear maintains an alternative world where the earth is flat and bleach eliminates Covid . Third, linear time is driven by Elon 's materialistic technology. Musk . AI is a fraud accelerator that thrives in the interstices of these temporalities.

# Recommendations

## 13 – The Controv3rse solution

Sam Altman, founder of Open AI, wants to have "new business models that will allow creators to generate revenue."

We have seen that the Uber model of data emerges to avoid the degeneration of generative AI, but that it threatens human capital. The economic model must rely on OGCs to open up a perspective of revaluation of this capital through liquid property and sacemization .

The business model we recommend to Sam Altman, and other generative AI operators, is this:

### Controv3rse Solution

RECOMMENDATION	OPEN AI	OGC
1 - Reserve	Reserve 15% of generative AI turnover for fair remuneration of authors and rights holders	Provide authors with a platform to object, where appropriate, to the use of their works by generative AI, or to choose a license, work by work
2 – Segmentation	Identify the main OGC by market segment (media/ dataset * country)	Claim a market segment (media/ dataset * country)
3 - Method	Publish the attribution method by OGC, detail the disambiguation in case of recovery	Negotiate coordination or reciprocity agreements between OGCs
4 – Non-Attribution	Identify a fund by country for unallocated amounts	Claim management of the fund
5 - Reports	Publish the summary in sufficient detail  Send to each identified OGC a monthly report containing the identifiers of the authors, the details of the use of their works, by synthetic content, and the value attributed	Send the operator the list of works that cannot be used by generative AI
6 - Payment	Distribute the reserve to OGCs and funds	Distribute the amounts collected to authors and rights holders

## 14 – Start the negotiation

### Steps

To reach agreements between generative AI operators and CMOs, we recommend that stakeholders document the following steps:

#### Steps of the Controv3rse solution

RECOMMENDATION	PROCESS	BEST PRACTICES, COMPARABLE
1 - Prototype	Discuss from a technical solution to be perfected to facilitate the projection of the interlocutors into a result, and reduce the complexity of the debate (see our demonstrator)	Demonstrator, trial and error methods
2 - Vision	Developing the Foresight: Decentralization vs. Centralized Platforms, Sacemization vs. Uberizing AI , Author Work vs. Commodification, Absurdity of Machine “Law,” Liquid Property, Utility Metrics	European framework: social market economy, economy of creativity, heritage and cultural diversity
3 - Fair remuneration	Specify the fair remuneration rate based on the measurement of the usefulness of all authors' works for all generated content. We propose a starting value of 15%.  Propose indicators of the usefulness of contributions for author-by-author attribution, and the method of refining this attribution in anticipation of future technical progress.	Spre Directive 2006/115/EC (13) Neighboring rights (73)  Yacast statistics for Spré
4 – Code of Good Practice by the AI Office	State of the art observatory <ul style="list-style-type: none"> <li>- Methods: Shapley value, pay -to-train, similarity</li> <li>- Arbitrations: parameters, granularity, utility indicators</li> <li>- Option and licensing platforms</li> <li>- Identification of rights holders and authors</li> <li>- Processing of unallocated amounts</li> </ul> <p>Sufficiently detailed summary:</p> <ul style="list-style-type: none"> <li>- Contents</li> <li>- Granularity</li> <li>- Format</li> </ul> <p>Reports to rights holders:</p> <ul style="list-style-type: none"> <li>- Dataset claims by rights holders</li> <li>- Methods of allocation and distribution by rights holder</li> <li>- Processing of recoveries</li> <li>- Data</li> </ul> <p>Secret Content: Information not disclosed in summaries and reports</p> <p>Progressive normalization</p>	Code of good practice against disinformation  Authors ' Guild Recommendations  Mission of the CSPLA  ISWC, ISAN
5 – Patents	Ensuring French and European positions through a portfolio of dedicated patents, in public-private cooperation	INPI Alliance PI Program



## Fair pay

Fair remuneration through generative AI could be inspired by that of the Spré . In France, this organization collects a royalty calculated on the turnover of nightclubs for the benefit of musicians. A representative sample of 120 of them is equipped with a " Yacast " box, financed by Sacem and Spré . Yacast collects the lists of music broadcast, and publishes statistics by artist, label, genre and number of weeks of broadcast. After a statutory deduction of 9%, Spré distributes the levy to the OGCs (SCPA 50%, ADAMI 25% and SPEDIDAM 25%) and communicates the Yacast statistics to them to remunerate the ultimate beneficiaries (performing artists, producers, general interest actions).

## Fair rate

In his 2020 report, *The Author and the Act of Creation*, Bruno Racine deplored the fact that "the proportional remuneration of authors only exceptionally reaches or exceeds 10% of operating revenues". He attributed this observation to "the relationship between the artist-author and downstream stakeholders (publishers, distributors, producers) [which] appears profoundly unbalanced, which leads, in his case, to calling into question in many cases the very idea of contractual freedom". To remedy this, he wanted to "strengthen artist-authors collectively, by quickly organizing professional elections that will give substance and legitimacy to the National Council of Artist-Authors to be created in order to serve as a framework for collective negotiation with distributors" to establish "a contractual system remunerating the author's working time, preempted by the sponsor".

Since the idea of revaluing labor value is not applicable to the case of generative AI, we propose to define the fair rate as the value of the utility of all the authors' contributions reported to the turnover of the generated content. It is on this principle that the Shapley value is designed and that the economist Ernst Fehr estimates at 14% the value brought by the press to Google Search . Since the famous question "was this comment useful to you ? " launched by Amazon, all commercial websites improve the quality of their offer by multiplying the utility metrics, which are now in common use.

Setting the basis on the turnover of the content generated would allow the fair remuneration of the author not to depend on the greater or lesser vertical segmentation of the sector, for example: author / platform / dataset / general model / fine tuning / interface / generation.

# 15 – License – Acknowledgments

This note and its tables are accessible under CC-BY license, subject to the fair sharing of the fruits of its exploitation by generative AI, granted by the association Controv3rse, RNA W751267044.

This document was automatically translated by Google Translate. In case of ambiguity, the French version prevails.

It draws on the reflections and discussions between members of the ministerial missions for the AI Act , those of the independent think tanks Controv3rse and Institut de l' Economie , as well as other figures recognized for their expertise. It extends the study previously carried out for the Pradal -Rambaud parliamentary mission and the columns published in *Le Monde* and *Les Echos* . Vincent Lorphelin, Founder

and co-President of these think tanks, author of this note, warmly thanks them for their ideas and the quality of these discussions

#### Presidents and Rapporteurs of the parliamentary and ministerial missions:

Alexandra Bensamoun , professor of law at the University of Paris Saclay, former member of the Artificial Intelligence Commission

Joëlle Farchy , university professor, Director of m2ecn and the PcEn Chair  
University Paris 1 Panthéon-Sorbonne, member of the CSPLA

Philippe Pradal , Member of Parliament for the Third Constituency of the Alpes Maritimes, National Assembly

Stéphane Rambaud, Member of Parliament for the Third Constituency of Var, National Assembly

Lionel Ferreira, Master of Requests at the Council of State

Julie Groffe -Charrier, HDR lecturer in private law at the Jean-Monnet Faculty

Bastien Blain, Junior Professor of Economics, Data and Behavioral Scientist, University of Paris 1 Panthéon-Sorbonne

#### Experts and Entrepreneurs:

William Bailey  
Co-Founder and President, Bolero Music

Emmanuel Benazera  
Chairman and CEO, Jolibrain

Jean-Paul Betbèze  
Economist , Betbèze Conseil

Michelle Bergadaà  
President of the Institute for Research and  
Action on Academic Fraud and Plagiarism

Michel Bokobza  
Mission Manager, Collège de Paris

Laurent Cervoni  
Director of Research and Innovation,  
Talan

Charles-Eric de la Chapelle  
Founder, Myriad Data

Jean-Philippe Clair  
Marketing, Communication and  
Innovation Director, Keyrus

Alice Coatalem  
Co-Founder of CogNeed and Associate  
Professor, Paris Dauphine University PSL

Christophe Collet  
Founder, AskLocala

Frederic Dayan  
Founder, Exactcure

Guillaume Desveaux  
Co-Founder of Aleia and Administrator,  
AI Cargo Foundation

Jonathan Dory  
Founder , Live Crew

William Eldin  
Founder, XXII

Stephanie Flacher  
Co-Founder, Logion Network

Raphael Frisch  
Co-Founder, HawAI.tech

Alain Garnier  
CEO, Jamespot

Anthony Graveline  
Founder, Disaitek

Philippe Guguen  
President, Sorga - Map Emulsion

Francis Hachem  
Founder, Codenekt

Rodolphe Hasselvander  
Founder, Blue Frog Robotics

Brice Hoarau  
Founder, Semdee

Matthias Houllier  
Co-Founder, Wintics

Casey Joly  
Lawyer, IPso

Jean Latger  
Founder, Oktal -SE

Frédéric Lefebvre- Naré  
AI Data Director, Niji

Youness Lemrabet  
Founder, Everysens

Jacques Levy- Vehel  
President of Case Law Analytics , and  
Director of Research, INRIA

Olivier Laborde  
Innovation and Digital Transformation  
Leader, BPCE

Alexandre Leforestier  
Founder, Panodysey

Sixtine Lorphelin  
AI engineer, UTC, student at INSEAD

Aymeric Masurelle  
Co- Founder , Spoon.ai

Pierre Miralles  
Co-Founder, Footovision

Emmanuel Moyrand  
Co-Founder, France Meta

Clement Merville  
Founder, Manzalab - Teemew

Edward of Miollis  
Founder, Polycube

Nathalie Nevejans  
Professor of private law and artificial  
intelligence, University of Artois

Remy Ozcan  
Co-Founder, Crypto4All

Jean-Jacques Quisquater  
Professor of Cryptography, Polytechnique  
Louvain, columnist for Le Monde

Jean-Michel Salomon  
President, Society of Game Authors

Frederic Soufflet  
Co-Founder, Haapic

Clement Tequi  
Co-Founder, Terno

François-Xavier Thoorens  
CEO, Vaultys

Arnaud Touati  
Associate Lawyer, #Hashtag

Christophe Tricot  
Co-Founder, La Forge

Killian Vermersch  
Co-Founder and CEO, Golem.ai

# Die Controverse-Lösung

Ein Wirtschaftsmodell zur Steigerung des kulturellen und menschlichen Kapitals durch



Adobe Firefly

C O N T R O V E R S E

# Zusammen- fassung

Controv3rse ist ein unabhängiger Think Tank, der von 70 Unternehmern und Experten für generative KI gegründet wurde. Dieser Bericht bietet eine Lösung als Antwort auf die Aussage von Sam Altman, dem Gründer von OpenAI, der sich für ein neues Wirtschaftsmodell einsetzt, das Entwickler angesichts des Aufstiegs der generativen KI belohnt.

Man muss sagen, dass der Konflikt zwischen den Autoren, deren Werke geplündert werden, und den Ausbeutern generativer KI, die Klarheit und Rechtssicherheit brauchen, vor Gericht mindestens zehn Jahre lang nicht gelöst werden wird.

Gleichzeitig verwässert die Verbreitung generativer KI die Originalität und löscht Spuren der Originalwerke aus. Gleichzeitig reagieren Micro-Tasking-Plattformen auf die enorme Nachfrage nach Daten, zum Nachteil von Modellen, die auf Kreativität und Urheberrecht basieren, wie etwa SACEM. Das kulturelle und menschliche Kapital, einst ein zentrales Element der europäischen Strategie, ist nun doppelt vom Raubbau bedroht.

Der Wirtschaftsnobelpreisträger Jean Tirole betont jedoch, dass nationaler Wohlstand auf der Förderung von Innovationen beruht, unterstützt durch geistiges Eigentum als Hebel für Kreativität. Um das kulturelle und menschliche Kapital zu bewahren und gleichzeitig Innovationen anzuregen, muss das neue Wirtschaftsmodell die Urhebergesellschaften neu erfinden.

Lloyd Shapley führte 1953 ein Schlüsselkonzept für die gerechte Verteilung von Gewinnen in einem kooperativen Spiel ein, das auf dem Beitrag jedes Spielers basiert. Die abgeleitete Methode wäre ideal, um die Autoren generativer KI zu vergüten, ist jedoch leider mit unerschwinglichen Rechenkosten verbunden. Es gibt Alternativen, wie beispielsweise das von Bilddatenbanken (Shutterstock, Getty) verwendete „Pay-to-Train“-Modell. Sie entschädigen die Autoren auf Grundlage der Quantität der bereitgestellten Inhalte, ignorieren jedoch die Qualität oder Nützlichkeit der Werke. Startups und Forscher ergänzen dieses Modell, um eine gerechtere Verteilung auf der Grundlage von Ähnlichkeiten zu gewährleisten. Wir bieten dem Leser einen Online-Simulator, mit dem er sich im realen Maßstab selbst von der technischen und wirtschaftlichen Machbarkeit dieses Ansatzes überzeugen kann.

Perspektivisch schützt dieses Modell die Rechte der Autoren und widersetzt sich der „Uberisierung der KI“. Die „Sacemisierung“ wertschätzt das „liquide Eigentum“, den intellektuellen Reichtum der Kreativität, der das kulturelle und menschliche Kapital angesichts seiner Ausbeutung durch Monopole erhält und entwickelt.

Wir empfehlen daher Sam Altman und den Betreibern generativer KI ein ökonomisches Modell in Zusammenarbeit mit den Autorengesellschaften und allen Beteiligten eine Methode, um Verhandlungen in diesem neuen Reflexionsrahmen anzustoßen.

# Heraus- forderungen



# 1 – Wer möchte, dass generative KI Autoren bezahlt?

„Ich denke, wir brauchen eine neue Vereinbarung, einen neuen Standard, ein neues Protokoll – wie auch immer Sie es nennen wollen – für die Art und Weise, wie Entwickler belohnt werden“, sagte Sam Altman, Gründer von „Open AI“. „Wir brauchen neue Wirtschaftsmodelle, die es den Kreativen ermöglichen, Einnahmen zu erzielen. »

Die Idee, Autoren durch Gen-KI zu bezahlen, wird seit mehreren Monaten auch von anderen Persönlichkeiten vertreten.

KI-Betreiber, die Autoren bezahlen wollen

NAME	UNTERNEHMEN	ZITAT
Sam Altman	Open AI	„Ich denke, wir brauchen eine neue Vereinbarung, einen neuen Standard, ein neues Protokoll, wie auch immer Sie es nennen wollen, um festzulegen, wie die Urheber belohnt werden. Wir brauchen neue Wirtschaftsmodelle, die es den Kreativen ermöglichen, Einnahmen zu erzielen. »
Arthur Mensch	Mistral AI	„Wir haben ein Interesse daran, Zugang zu qualitativ hochwertigen Inhalten zu haben. Wir haben daher Gespräche mit Inhabern, im Verlagswesen und in der Presse aufgenommen. Es können Synergien und Mechanismen zur Wertteilung gefunden werden. »
David Holz	Midjourney	„Es gibt keine wirkliche Möglichkeit, hundert Millionen Bilder zu nehmen und herauszufinden, woher sie stammen. Es wäre schön, wenn die Bilder eingebettete Metadaten über den Urheberrechtsinhaber oder etwas Ähnliches hätten. Dies ist jedoch nicht möglich, da kein Register vorhanden ist. Es gibt keine Möglichkeit, ein Bild im Internet zu finden und es dann automatisch seinem Besitzer zuzuordnen und anschließend irgendetwas zu seiner Authentifizierung zu unternehmen. »
Chris Dixon	a16z	„Die [Web 3]-Bewegung hat das Potenzial, den Geist des frühen Internets zurückzubringen, die Eigentumsrechte der Urheber zu sichern, den Nutzern Eigentumsrechte und Kontrolle zurückzugeben und den Würgegriff der Großkonzerne zu brechen, die unser Leben zentral beherrschen. ".
Neal Stephenson	Lamina1 (Erfinder des Metaverse)	„Sie können tatsächlich verfolgen, wie viel Einfluss ein bestimmtes Bild oder eine bestimmte Eingabe auf das Endergebnis hatte. Und wenn das Endergebnis dann etwas wert ist, dann gibt es vielleicht eine Möglichkeit, diesen Input anzuerkennen und angemessen zu vergüten [...] Nehmen wir an, ich schreibe ein Buch, in dem es um ein magisches Schwert geht, das gerade im Buch beschrieben wird. [...] Jemand könnte dann ein Asset erstellen, das im Unreal Asset Store verkauft werden könnte. [...] Es gibt Sounddesigner, die etwas Ähnliches mit den Geräuschen machen müssen, die bei der Verwendung entstehen. Und es gibt Programmierer, die das Schwert mithilfe von Blaupausen oder C++ in das Spiel integrieren müssen, damit es tatsächlich etwas tun und zum Spielerlebnis beitragen kann. So wird mit jedem Schritt mehr Wert geschaffen. Und am Ende dieses Prozesses haben Sie etwas, das tatsächlich Einnahmen generieren könnte. Und wenn die Einnahmen dann endlich eingehen, soll sich diese Entwicklung ja auch in die umgekehrte Richtung auswirken. Und Sie möchten, dass die verschiedenen Menschen, die zur Wertschöpfungskette beigetragen haben, in irgendeiner Weise entschädigt werden.“



NAME	UNTERNEHMEN	ZITAT
Jaron Lanier	Microsoft	„Die Leute werden für ihre Daten bezahlt und werden für Dienste bezahlen, die Daten anderer Leute erfordern. Die Aufmerksamkeit des Einzelnen wird von seinen selbst definierten Interessen gelenkt und nicht von manipulativen Plattformen, die Werbetreibenden oder anderen Drittparteien verpflichtet sind. Plattformen erhalten qualitativ hochwertigere Daten, mit denen sie ihre maschinellen Lernsysteme trainieren können. » Lanier räumt ein, dass sich selbst Forscher, die sich mit der Würde der Daten beschäftigen, nicht darauf einigen können, wie man alles entwirrt, was die KI-Modelle aufgenommen haben, oder wie viele Details nötig sind, um dies zu erklären. Lanier glaubt jedoch, dass es möglich sei – schrittweise. „Wir müssen die Herkunft der menschlichen Quellen berechnen und darstellen, die für ein bestimmtes KI-Ergebnis am wichtigsten waren. Das machen wir derzeit nicht. Wir können es effektiv und effizient tun, wir tun es einfach nicht. Es muss eine gesellschaftliche Entscheidung sein, diesen Schritt zu gehen.“
Yat Siu	Animoca-Brand	„Daten sind die wertvollste Ressource. Zunächst einmal sind sie die Grundlage für die künstliche Intelligenz, von der wir so viel hören. Keine Daten, kein ChatGPT . Daten sind die neue Arbeit. Und wir werden dafür nicht angemessen entlohnt. [...] Es besteht ein klarer Zusammenhang zwischen Eigentumsrechten und dem Wohlstand einer Nation. » „Stellen Sie sich vor, Sie gehen auf Facebook und stellen am Ende des Tages fest, dass Sie gerade tausend Dollar für das Unternehmen verdient haben. Ihre Beziehung zu Facebook würde sich völlig ändern. Sie würden wahrscheinlich Ihren gerechten Anteil verlangen. »
Yann Le Cun	Meta	„ein Modell zur Verbreitung von [Urheberrechts-]Rechten in die eine und von Einnahmen in die andere Richtung“
Reise Adler	ehemaliger CEO von Scribd	„Das vierte Gesetz ist eine Reihe von Leitprinzipien dafür, wie KI-Unternehmen von Menschen erstellte Inhalte nutzen und damit trainieren können. Das vierte Gesetz, das von den drei Robotergesetzen des Science-Fiction-Autors Isaac Asimov inspiriert wurde, besagt, dass Menschen das Recht haben sollten, der Nutzung ihrer Arbeit durch KI zuzustimmen und diese zu kontrollieren, und dass sie (falls gewünscht) für ihre Arbeit entschädigt und anerkannt werden sollten (wenn ein Buch im Ergebnis referenziert wird, sollte ein Link zum Kauf vorhanden sein). »

Unter der Führung von Sam Altman erhält die Idee einer Entschädigung der Autoren mehr Substanz, obwohl er der Ansicht ist, dass eine solche Entschädigung nur dann fällig wäre, wenn der Stil des Autors plagiariert würde, was sehr einschränkend sei.

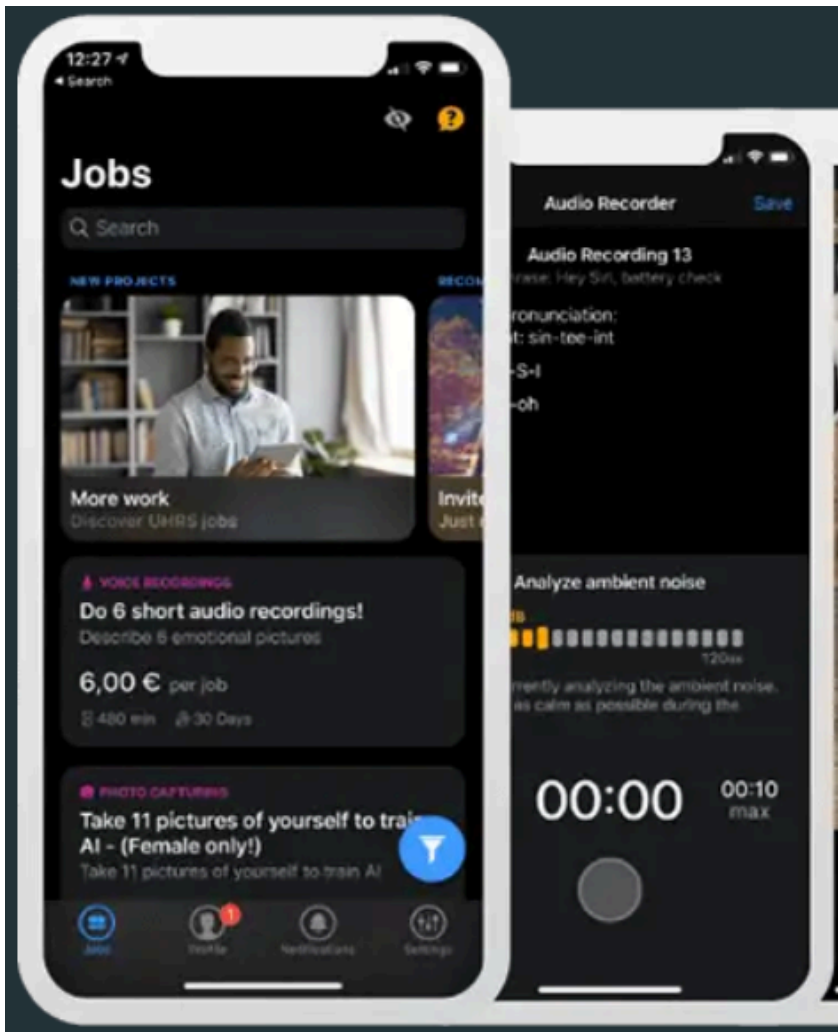
Er stellt jedoch das wirtschaftliche Modell seiner Umsetzung in Frage . Der Zweck dieser Notiz besteht darin, dieses neue Wirtschaftsmodell im Rahmen des KI -Gesetzes vorzuschlagen .

## 2 – Die drei Gesichter der Autorenvergütung

das Scraping , der die Urheber gegen die Betreiber generativer KI ausspielt, wird, würde er auf rein juristischer Ebene ausgetragen, noch mindestens zehn Jahre dauern.

In der Zwischenzeit wird sich die Branche in Datenersteller, -plattformen, -aggregatoren, -verhandler, -trainer, -veredler, -ausnutzer und -integratoren generativer KI segmentieren, was den Schöpfungsakt verwässern wird. Das kulturelle Erbe wird vermischt, verbreitet und recycelt, wobei die ursprünglichen Spuren des Inhalts ausgelöscht sein werden.

Aus Angst vor einer Degeneration der Lerninhalte entwickelt sich bereits ein neuer Sektor der Microtasks zur Massenproduktion von Inhalten.



©Clickworker 2024

Gig- Economy- Plattformen für KI basieren auf Befehlslogik. Clickworker zahlt Nutzern direkt 6 € dafür, dass sie 6 Bilder laut beschreiben und die Aufnahme veröffentlichen. Oder um ein Foto von sich selbst bei körperlichen Übungen zu machen. Gigwalk und Premise zahlen 4 \$, um ein Foto von einem Verkehrsstau an einer Kreuzung zu machen. Das Ergebnis wird zum Trainieren von KIs verwendet.

Dieses Modell vom Typ „Uber der Daten“ wächst schnell dank des enormen Bedarfs an Daten von Betreibern generativer KI.

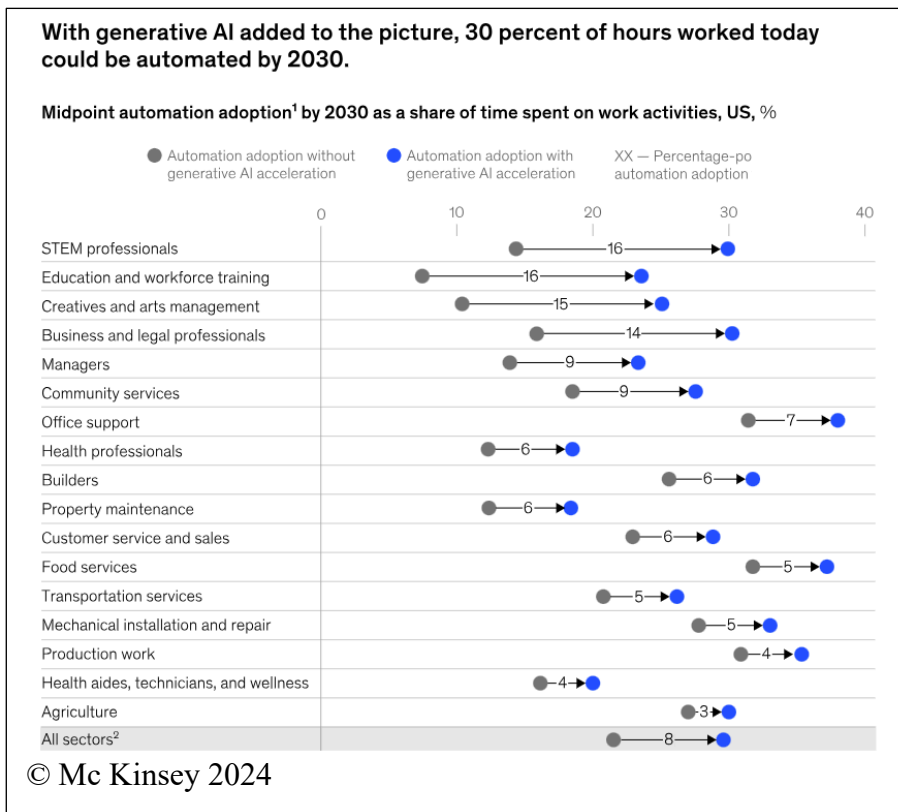
Dies ist das Gegenteil von Urheberrechtsgesellschaften wie SACEM, die auf dem Angebot und dem geistigen Eigentum der Autoren basieren, dieses Eigentum verfolgen, Lizenzgebühren einziehen und Rechteinhaber nach vorab festgelegten

Verteilungsschlüsseln vergüten.

Das erste Modell bevorzugt Nachfrage, Ausführung, sofortige Kommerzialisierung, kognitiven Neo-Taylorismus und die Standardisierung der Arbeit. Die zweite fördert Angebot, Initiative, geistiges Eigentum, Kreativität und kulturelle Vielfalt. Die einen nutzen kulturelles Kapital, die anderen pflegen es.

Neben anderen Wirtschaftsmodellen (siehe unten) sind Scraping, Daten-Uber und Autorengesellschaften die drei wichtigsten Gesichter der Vergütung – oder auch Nichtvergütung – von Autoren, deren Ungleichgewicht die aktuellen Spannungen verschärft.

### 3 – Über das Urheberrecht hinaus: Humankapital



Angesichts der Entwicklung generativer KI stehen Autoren und Kreativität an vorderster Front. Danach folgen das Handwerk, die künstlerischen Berufe, das Gutachten, die Forschung und alle Berufe mit Know-how in allen Wirtschaftsbereichen.

Überisierung der KI entstehen, so wie das Unternehmen Uber einst das Phänomen der Überisierung initiierte.

Wenn Sie beispielsweise die generative KI des Unternehmens mit all ihren Datenarchiven anpassen, können Sie für jede neue Nachricht, jeden neuen Bericht, jedes neue Projekt, jede neue Präsentation und jede neue Preisgestaltung einen Entwurf

erhalten. Durch professionelle Korrekturen wird die KI solange trainiert, bis die Fehlerquote akzeptabel wird. Die Produktivitätsgewinne sind erheblich, doch im Gegenzug schwächen sich die Positionen der Unternehmen an den strategischen Gliedern der Wertschöpfung zugunsten der KI-Betreiber.

Da all diese Fähigkeiten die Schnittstelle zwischen Wissen, Praktiken und sozialen Organisationsmaterialien sind, steht das Gefüge der Unternehmen auf dem Spiel. Es ist, als ob zur Zeit der industriellen Revolution ein Kartell von Besitzern der Dampfmaschinen die Kontrolle über die tayloristische Arbeitsorganisation. Dies sind die Erwartungen der Börse an die großen Technologieunternehmen. Durch die Gewährung von Billionenkaptalisierungen werden nicht mehr Unternehmen, sondern Monopole bewertet.

Die Vorstellung, dass Maschinen das „Recht haben, kostenlos zu lernen“, wird sich verbreiten. Selbst wenn die Urheber ihren Rechtsstreit gegen das Scraping in zehn Jahren gewinnen sollten, wird ihre Entschädigung auf den von den Uber-Daten bereitgestellten Vergleichswerten basieren und daher nicht viel betragen. Im Rückblick werden wir erkennen, dass das bestehende Urheberrecht nichts weiter als eine Maginot-Linie war, die leicht zu umgehen war.

Die Wirtschaft verfügt über kulturelles und Humankapital, das seit langem als erheblicher Vermögenswert angesehen wird und sogar in den Mittelpunkt der europäischen Wirtschaftsstrategie gerückt ist. Diese zu frühe Strategie geriet mangels eines ökonomischen Modells letztlich in Vergessenheit. Dieses unterbewertete und schlecht verteidigte Kapital ist nun durch die Überisierung der KI von einer Ausbeutung bedroht. Für jedes Unternehmen, jedes Land und jeden Kontinent ist die Ablehnung von Technologie

gleichbedeutend mit der Verurteilung zu einem dauerhaften Niedergang, wie es der Fall Chinas über zwei Jahrhunderte hinweg war. Sie ohne Entschädigung zu akzeptieren, hieße, sich kampflos kolonisieren zu lassen.

Der Nobelpreisträger Jean Tirole lehrt uns, dass 1/ der Wohlstand einer Nation von ihrer Fähigkeit abhängt, Wert auf der Ebene der Innovation zu schaffen. 2/ Innovation entsteht aus individueller Initiative, denn das Charakteristische an kreativer Arbeit ist, dass wir nicht wissen, was wir finden werden. 3/ Geistiges Eigentum ist ein notwendiges Übel, um diese Kreativität zu fördern. 4/ Einem Unternehmen die Vermarktung seiner Produkte zu untersagen, bis es seine Lizenzgebühren bezahlt hat, ist eine sehr wirksame Waffe gegen Gafam . 5/ Pools für geistiges Eigentum mit individueller Lizenzierungsberechtigung würden von ihnen verlangen, Lizenzen zu erwerben, allerdings zu einem wettbewerbsfähigen Preis.

Nationen, die ihr Human- und Kulturkapital steigern wollen, müssen eine moderne Sacemizing- Lösung finden , deren Wirtschaftsmodell wirksam mit der Uberisierung der KI konkurrieren kann .

# Technische Lösung

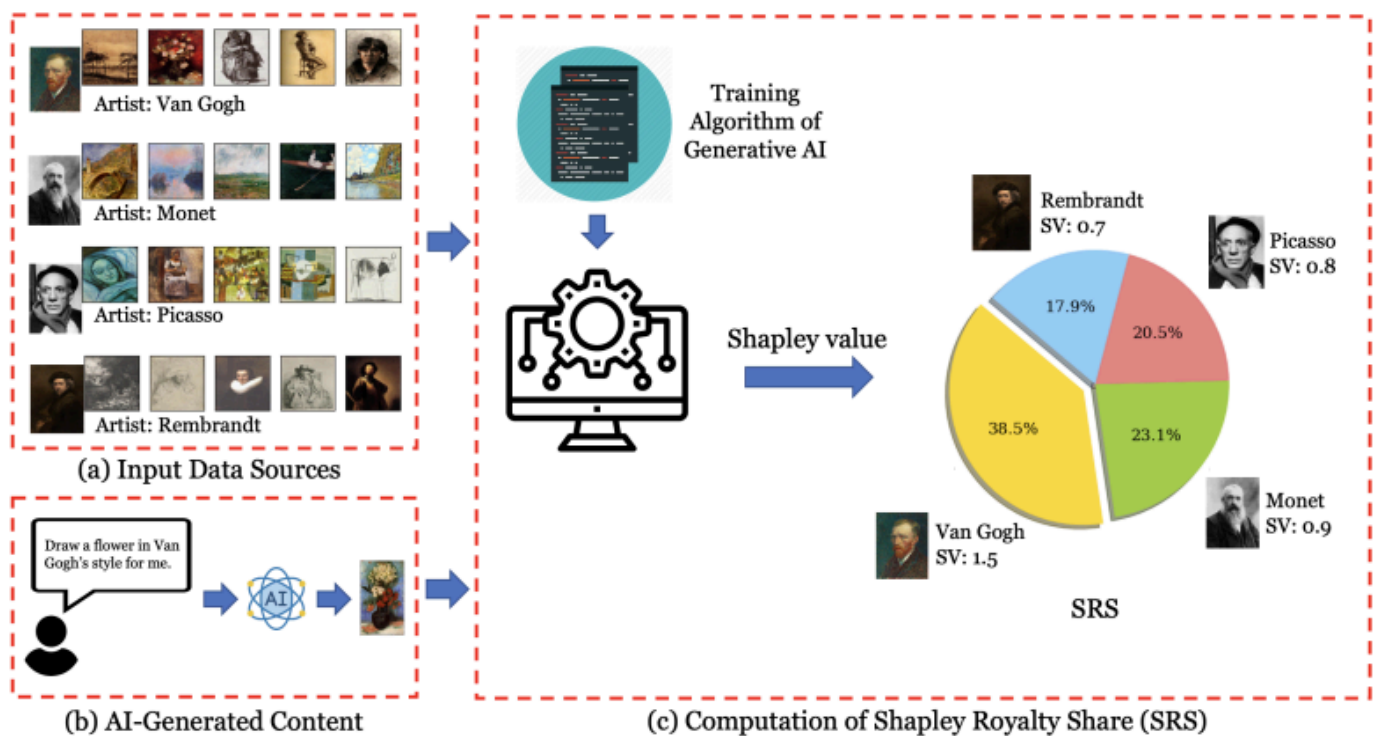
## 4 - Ideale Lösung

In der Spieltheorie führte Lloyd Shapley 1953 die Idee ein, dass es in einem kooperativen Spiel möglich sei, eine gerechte Verteilung der Gewinne unter den Spielern zu berechnen. Wir betrachten beispielsweise ein Meeresökosystem mit all seinen Lebewesen – Algen, Anemonen, Muscheln, Seesterne – und messen den Nutzen jeder einzelnen Art anhand ihres Beitrags zum Leben im Ökosystem. Wenn man Seesterne entfernt, verschwinden 50 % aller anderen Arten, im Vergleich zu 10 % bei der Entfernung von Anemonen. Der Shapley-Wert von Seesternen ist fünfmal so hoch wie der von Anemonen.

Über die ideale Lösung zur Vergütung der Autoren besteht in der Forschung Konsens. Dabei wird für jeden Autor der Shapley-Wert berechnet und die Vergütung anteilig zu diesem Wert verteilt.

Bei generativer KI sollte die gerechte Verteilung der Vergütung für Trainingsarbeiten idealerweise auf Basis der Nützlichkeit der Trainingsarbeiten für die KI-generierten Inhalte je Autor erfolgen. In der folgenden Abbildung beträgt Van Goghs Shapley-Wert 1,5 oder 38,5 % der Gesamtvergütung.

Shapley- Wert



In der Praxis sind die Kosten für die Berechnung des Shapley-Wertes unerschwinglich, was die Frage nach der technischen Machbarkeit einer fairen Vergütungslösung aufwirft.

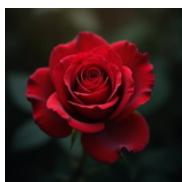
## 5 – Demonstrator

### Objekt

Der Leser findet einen Online-Simulator unter [https://huggingface.co/spaces/TheFrenchDemos/genai-rew\\_a rd](https://huggingface.co/spaces/TheFrenchDemos/genai-rew_a rd)







Wir glauben, dass es mittlerweile technisch und wirtschaftlich möglich ist, einen Attributionsscore für Schulungsinhalte zu berechnen, um eine gerechte Verteilung der Vergütungen zwischen den Rechteinhabern dieser Inhalte zu erreichen. Mit diesem Simulator kann jeder seine eigenen wirtschaftlichen Hypothesen selbst überprüfen.

Beispielsweise wurde mit der Eingabeaufforderung „*Eine leuchtend rote Rose in voller Blüte, Makrofotografie*“ ein Bild einer Rose generiert :



Der Simulator identifiziert in wenigen Sekunden unter den zehn Millionen Inhalten der Open Images-Datenbank die acht Originalbilder, die am meisten zur Entstehung des generierten Inhalts beigetragen haben. Es berechnet einen Zuordnungssatz und eine Vergütung und stellt Identifikationsinformationen über den Originalinhalt und seinen Autor bereit.

#### Zuschreibungen von Rosenbildern

INHALT	ZUSCHREIBUNG	INHALTSBELOHNUNG (c€)	Inhalts-URL	AUTORENNAME	AUTOR-URL
	21 %	0,63	<a href="https://www.flickr.com/photos/50093642@N03/4629227563">https://www.flickr.com/photos/50093642@N03/4629227563</a>	Yoko Nekonomania	<a href="https://www.flickr.com/photos/nekonomania/">https://www.flickr.com/photos/nekonomania/</a>
	18 %	0,54	<a href="https://www.flickr.com/photos/22305657@N02/11733854773tt">https://www.flickr.com/photos/22305657@N02/11733854773tt</a>	Bruno Pinho	<a href="https://www.flickr.com/photos/brunopinho/">https://www.flickr.com/photos/brunopinho/</a>
	15 %	0,45	<a href="https://www.flickr.com/photos/12692992@N00/170278141">https://www.flickr.com/photos/12692992@N00/170278141</a>	Dean + Barb	<a href="https://www.flickr.com/photos/barbandean/">https://www.flickr.com/photos/barbandean/</a>
	12 %	0,36	<a href="https://www.flickr.com/photos/33140999@N07/3509399376">https://www.flickr.com/photos/33140999@N07/3509399376</a>	Michael Arrighi	<a href="https://www.flickr.com/photos/arrighi/">https://www.flickr.com/photos/arrighi/</a>
	10 %	0,30	<a href="https://www.flickr.com/photos/55658968@N00/8730223835">https://www.flickr.com/photos/55658968@N00/8730223835</a>	Toshihiro Gamo	<a href="https://www.flickr.com/photos/dakiny/">https://www.flickr.com/photos/dakiny/</a>
	9 %	0,27	<a href="https://www.flickr.com/photos/34226407@N04/8734050294">https://www.flickr.com/photos/34226407@N04/8734050294</a>	Edd Prince	<a href="https://www.flickr.com/photos/edpics/">https://www.flickr.com/photos/edpics/</a>

	9 %	0,27	<a href="https://www.flickr.com/photos/41353201@N07/4267426012">https://www.flickr.com/photos/41353201@N07/4267426012</a>	Musiker Loxton	<a href="https://www.flickr.com/photos/dj-dwayne/">https://www.flickr.com/photos/dj-dwayne/</a>
	8 %	0,24	<a href="https://www.flickr.com/photos/64574638@N04/6006728032">https://www.flickr.com/photos/64574638@N04/6006728032</a>	serbischstruhpro	<a href="https://www.flickr.com/photos/serbestruhpro/">https://www.flickr.com/photos/serbestruhpro/</a>

Unten auf der Webseite ermöglichen drei Schaltflächen dem Benutzer den Download:

- Dieser Bericht
- Der Bericht der Rechteinhaber, den der Betreiber im Rahmen von Vergleichsvereinbarungen und der Rückübertragung angemessener Vergütungen erstellen und den Rechteinhabern übermitteln könnte
- Der gesetzliche Bericht, der im AI- Gesetz vorgesehen ist (Art. 53, 1, d), gemäß der vom Hohen Rat für Literatur und Kunst vorgeschlagenen Form Eigentum (CSPLA)

## So funktioniert es

Der Simulator hat eine Schaltfläche " Erweitert " Parameter », mit dem der Benutzer die gewünschten Parameter anpassen kann. Die folgende Tabelle kopiert die standardmäßig eingestellten Werte und gibt die Berechnungsformeln an.

Parameter und Werte einer Simulation

VARIABLE	EINSTELLUNG	VERGLEICHBAR	BERECHNUNG	FORMEL
Benutzerabonnement	12 €/Monat (1)	Adobe Firefly ab 4,99 \$/Monat Midjourney ab 10 \$/Monat Dall-e in ChatGPT Plus enthalten \$20 Getty Edify 45 € für 25 Generationen	S	
Generierte Inhalte	60 /Monat	Adobe Firefly Basic Plan 100/Monat Midjourney Basisplan 100/Monat	G	
Nennwert pro generiertem Inhalt			0,2 €	$V = S / G$
Anteil der Rechteinhaber und Autoren				$R\% = Rr\% + Ra\%$
- Rechteinhaber	10 %	Musik (ASCAP, BMI, SACEM) 2 – 8 % Stockfoto 55 – 85%	Rr %	
- Autoren	5 %	Gedruckte Bücher 5 – 15 % Musik (Interpret und Komponist) 15 – 30 %	%	



## Stockfotografie 15 – 45 %

Vergütung der Rechteinhaber und Autoren pro generiertem Inhalt		3 c€	$R = V \times R \%$
Namensnennung		21 %	$A\% =$ Ähnlichkeitsalgorithmus
		8	N = Anzahl der zugewiesenen Inhalte
Belohnung für Trainingsinhalte		0,63 c€	$T = R \times A \%$
Anzahl der Autoren	1.800		B
Benutzer	500.000	Midjourney wird auf 2 bis 5 Millionen geschätzt .	U
Monatliche Belohnung für Rechteinhaber und Autor		829 €	$S \times U \times R\% / B \times A\% \times N$
- Rechteinhaber		276 €	$S \times U \times Rr\% / B \times A\% \times N$
- Autor		553 €	$S \times U \times Ra\% / B \times A\% \times N$

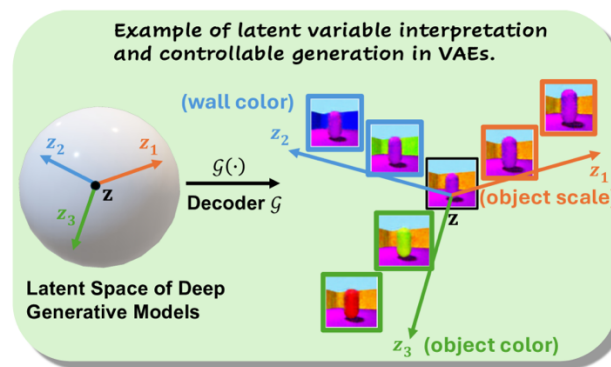
## Konzept der Ähnlichkeit

Dieser Demonstrator verwendet einen Ähnlichkeitsalgorithmus zwischen Trainingsinhalten und generierten Inhalten und implementiert [DINOv2](#) und [FAISS](#). Intuitiv sehen zwei ähnliche Bilder gleich aus.

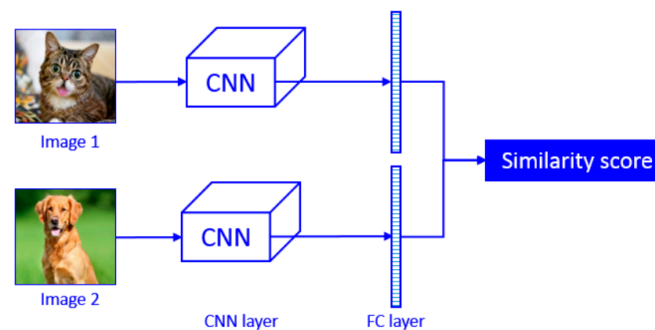
jedoch um einen wissenschaftlichen Begriff, der losgelöst ist von jedem juristischen Begriff der „Fälschung“, „Ähnlichkeit“ oder Verletzung eines Werks, das im französischen Recht als „jede Vervielfältigung, Darstellung oder Verbreitung mit welchen Mitteln auch immer“ definiert ist. Mitteln, eines geistigen Werks unter Verletzung der gesetzlich definierten und geregelten Urheberrechte“ (CPI, Art. L. 335-3). Eine Rechtsverletzung bzw. ein Verstoß liegt nicht zwangsläufig bei der Reproduktion eines Werkbestandteils, in diesem Fall eines „Beitrags“, vor. Auf die gleiche Weise kann bei der Erstellung eines Werks ein Element in die Komposition einfließen, ohne dass es zu Fälschungen oder Rechtsverletzungen durch Dritte kommt. Die aus der „Ähnlichkeit“ resultierende Vergütung dürfe den Urheber im Verletzungsfall daher nicht in seiner Klagefähigkeit behindern.

## Vektorähnlichkeit

Technisch gesehen wird jeder generierte oder ursprüngliche Inhalt analysiert, um messbare Informationen zu extrahieren: Katze, Hund, Vordergrundfarbe und -größe, Typ, Stil, Größe, erkannte Objekte, Schlüsselwörter, ... all diese Hunderte von Nutzenindikatoren , etwa 500 in unserem Demonstrator , bildet einen Vektor in einem „latenten Raum“.



Der Grad der Zuordnung eines Werks (Bild 1) zu einem generierten Inhalt (Bild 2) wird durch einen Ähnlichkeitswert zwischen den beiden Bildern berechnet, der von den beiden siamesischen neuronalen Netzwerken verarbeitet wird. Je näher die resultierenden Vektoren beieinander liegen, desto höher ist der Ähnlichkeitswert zwischen den Bildern.



## Anzahl der Autoren

Die Open Images-Datenbank wurde aus der Datenbank von Flickr gebildet und verwendete Filter wie Nichterkennung durch die Google Bildersuche, eine Mindestanzahl von Pixeln oder Deduplizierung. Auch Open Images wählt nur CC-BY-Lizenzen aus, also ein Bild unter 30. Da Flickr standardmäßig die Lizenz „Alle Rechte vorbehalten“ anbietet, dürfen Lizenzänderungen nur von informierten Autoren vorgenommen werden. Dies deutet auf eine hohe Konzentration der aktivsten Mitwirkenden bei Open Images hin, die mehr als 5.623 Fotos oder rund 1.800 Autoren veröffentlicht haben. Diese intuitive Hypothese könnte durch eine genaue Zählung der Autoren der Open Images-Datenbank verifiziert werden, was uns aufgrund der begrenzten Abfragegeschwindigkeit nicht möglich war.

## Identifizierung der Autoren

Die Open Images-Datenbank identifiziert die Autoren der Bilder und stellt ihre Flickr-Profilseite bereit, die jedoch keine vollständige E-Mail- oder Postadresse enthält.

Über unseren Demonstrator und die Open Images hinaus bieten öffentliche Datensätze Identifikationselemente mit unterschiedlicher Genauigkeit, die jedoch durch Querverweise identifiziert werden können, um jeden Rechteinhaber zu identifizieren.

Querverweismethoden sind vor allem für ihren betrügerischen Einsatz zum Schutz der Privatsphäre bekannt. Schon 1997 war es dank der Pflicht zur Übermittlung aggregierter Statistiken zusammen mit Wahldaten durch einen Versicherer aus Massachusetts möglich, Gouverneur William Weld fehlerfrei zu identifizieren und seine Gesundheitsgeschichte zu kennen.

Ihre legale Verwendung ermöglicht in den meisten Fällen die vollständige Identifizierung der Autoren: durch Vergleich sozialer Graphen, durch statistische Korrelation dank einer seltenen Kombination von Attributen (wissenschaftliche Nische, Aktivitätsstunden, IP-Adresse), durch Rekonstruktion einzelner Attribute aus Gruppendurchschnitten und bekannten Eigenschaften anderer Personen usw. Die folgende Tabelle bietet Hinweise zur Identifizierung der Autoren von GPT 4-Schulungsinhalten.



AI-generated by Adobe Firefly

Mögliche Identifizierung von Autoren der von GPT-4 verwendeten Datensätze

	DATENSATZ	IDENTIFIZIERUNGSELEMENTE	DIREKTE ÜBERGANGSSTUFE	KENNZEICHNUNG
<b>Direkte Identifizierung</b>	Bilder öffnen	-	-	Name, Vorname, Steckbrief
	Project Gutenberg	-	-	Name, Vorname, Geburts- und Sterbedaten
	arXiv	-	-	Name, Vorname_Initial
	PubMed	-	-	Id_Autor
<b>Indirekte Identifizierung</b>	Allgemein	Domain	Whois	Registrant_contact (Name, Organisation, Straße, Stadt, Staat, Telefon)
	BuchCorpus	Plain_text	Schlagwörter	Name, Vorname
	WebText	Plain_text + Suche = "geschrieben von"	Amazon API	Name, Vorname
	StackOverflow	HTML	Stackoverflow.com	Benutzer-ID , Netzwerkprofil
<b>In warten auf Anspruch</b>	Wikipedia	Benutzername	-	-
	Reddit	Benutzername	-	-
<b>Andere</b>	Gemeinsame Stimme	(Statistiken)	-	-

Google Bücher Ngrams	( nicht zutreffend )	( n / A )	( n / A )
GitHub	( nicht zugänglich )		

## 6 – Vorhandene Lösungen

### Synthese

In der folgenden Tabelle sind einige kommerzialisierte oder getestete Methoden in Bezug auf unseren Demonstrator aufgeführt:

Methoden zur Förderung von Autoreninhalten

BEWERTUNG	BEWERTUNGSMETHODE	DEMONSTRATOR	Handelsunternehmen	FORSCHER
Generierte Inhalte	Abonnementkosten/Menge	Kontroverse		
Datensatz	Ähnlichkeit von Inhaltssets Shapley Anlaufdaten Shapley Häufigkeit des Website-Zugriffs, Art der Daten		Tollbit	Adobe Offene KI Offene KI
Trainingsarbeit	Ähnlichkeit des Inhalts Modelliertes Lernen Vorlagenanpassung Verlernen Tätowieren Shapley föderierte	Kontroverse	Bria (Bilder), Prorata Gist.ai (Texte)	Offene KI Adobe Adobe Adobe Huawei
Autor	Bezahlen für die Ausbildung Einfluss des Gesamtwerkes Identifizierung der Rechteinhaber		Shutterstock , Getty , Audiosparx , Canva Sureel.ai (Musik)	
		Kontroverse		

## Vermarktete Lösungen

Unternehmen, die Autoren für Schulungsinhalte bezahlen

UNTERNEHMEN	RISIKOKAPITAL	UMSATZ	VERFAHREN	MEDIEN	INHALTSANBIETER
Adobe (Firefly)	-	21 Milliarden US-Dollar	Bezahlte Ausbildung	Bild	
Leinwand	-	2 Milliarden Dollar	Bezahlte Ausbildung	Bild	
Getty	-	917 Millionen US-Dollar	Bezahlte Ausbildung	Bild	
Shutterstock	-	875 Millionen US-Dollar	Bezahlte Ausbildung	Bild	
ProRata (Gist.ai)	25 Millionen US-Dollar	-	Inhaltsähnlichkeit	Text	Sky News, Time, Financial Times, The Atlantic, Axel Springer, The Guardian, dmg media, Universal Music, Fortune
Bria	24 Millionen US-Dollar	-	Inhaltsähnlichkeit	Bild	Shutterstock, Getty Images, Alamy, Superstock
Tollbit	24 Millionen US-Dollar	-	Häufigkeit des Zugriffs auf die Webseite	Text	Penske Media Corporation, Time, Mumsnet, Vertrauenswürdige Medienmarken, Candr Media Group, AdWeek
Sureel.ai	-	-	Einfluss des Gesamtwerkes	Bild, Musik	
Audiosparx	-	-	Bezahlte Ausbildung	Musik	

## Pay -to-train -Methode

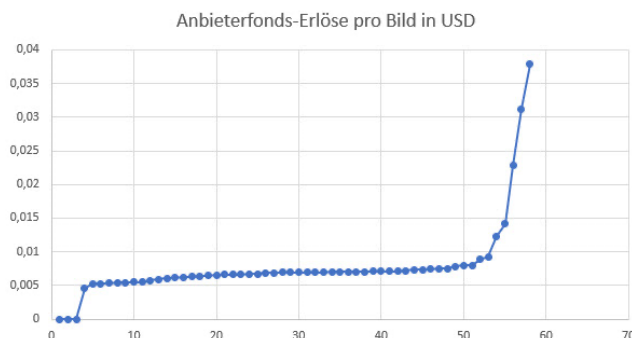


AI-generated by DALL-e

Die privaten Inhaltsbanken von Adobe, Canva , Getty und Shutterstock , verteilen den Preis der von ihnen verkauften Datensätze nach der Methode „ Pay to Train“. Die Autoren werden auf Basis des Umfangs ihrer Beiträge zu diesen Datensätzen vergütet .

Der „ Pay -to-Train“-Wert lässt sich aus verschiedenen Nutzenindikatoren berechnen. Shutterstock zahlt beispielsweise 0,0050 \$ pro Foto und Jahr, wobei je nach Menge der Metadaten und anderen Kennzahlen bis zu 0,0038 \$ pro Foto und Jahr möglich sind.

Pay -to-Train- Wert : Autorenvergütung durch Shutterstock . Stichprobe N = 58 (in \$/Bild)



Der Vorteil dieses Modells liegt in der einfachen Berechnung. Es ist seit zwei Jahren in Betrieb und scheint bei den Rechteinhabern gut angenommen zu werden.

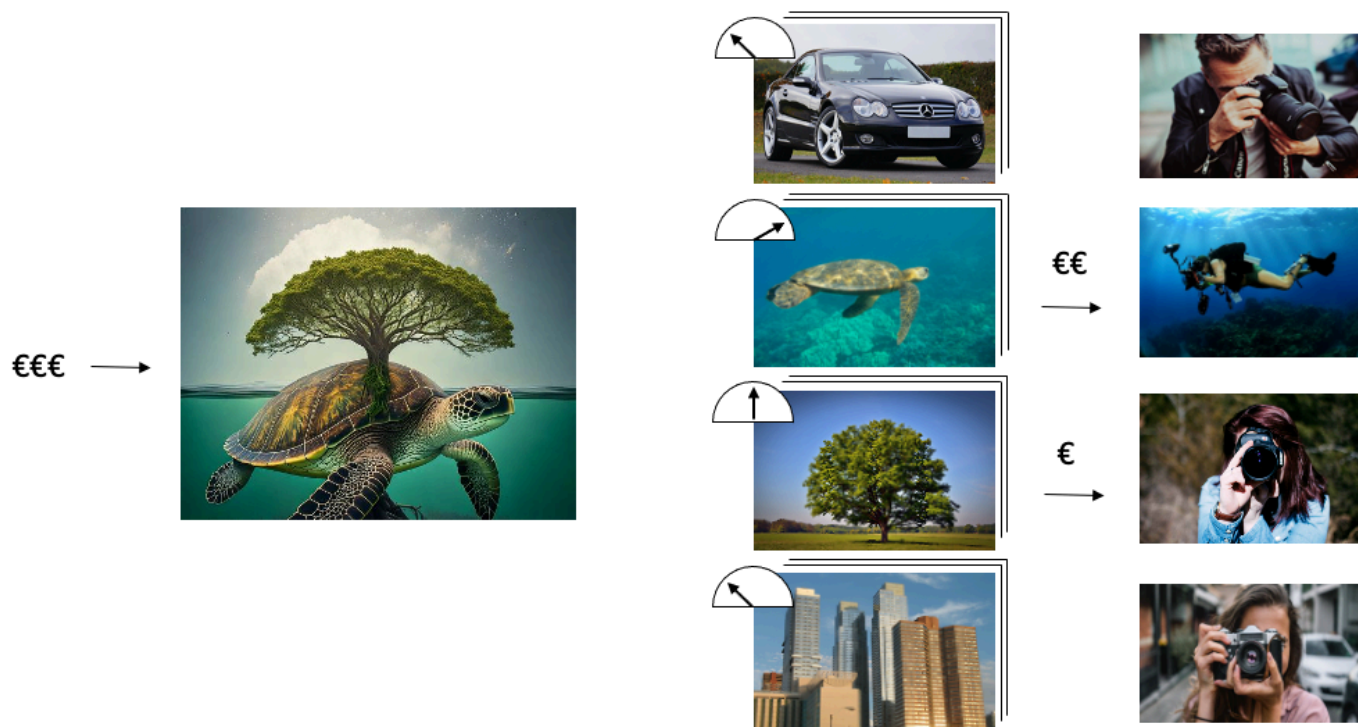
Andererseits gibt es hinsichtlich der Fairness zwischen den Autoren zwei Verzerrungen:

- (11) Die Qualität des Inhalts wird dabei nicht berücksichtigt: Ein Autor, der 200 Fotos beigesteuert hat, erhält bei sonst gleichen Bedingungen die doppelte Vergütung eines Mitwirkenden, der 100 Fotos beigesteuert hat.
- (12) Die Nützlichkeit der generierten Inhalte wird nicht berücksichtigt: Van Gogh erhält nicht mehr als Picasso, wenn Benutzer in ihren Eingabeaufforderungen öfter „in der Art von Van Gogh“ einfügen.

## Hybridmethode Ähnlichkeit + Pay -to-train

Das israelisch -amerikanische Startup Bria, das mit großen privaten Bilddatenbanken ( Shutterstock , Getty) zusammenarbeitet, wird von Big Tech (Microsoft, Amazon, Nvidia ) integriert . Als ersten Ansatz verwendet Bria, wie auch unser Demonstrator, ein Ähnlichkeitsverfahren, um die Autoren proportional zur so gemessenen Ähnlichkeit zu vergüten.

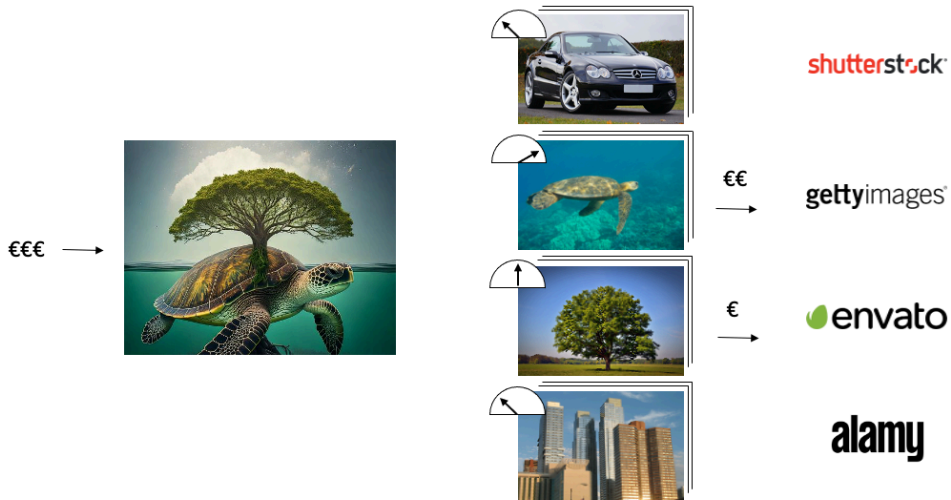
Messung der Ähnlichkeiten zwischen generierten Inhalten und Originalfotos zur Vergütung der Autoren



Genauer gesagt scheint es so zu sein, dass die Summe der Ähnlichkeitsbewertungen von Bilddatenbanken ( Shutterstock , Getty, Alamy usw.) berechnet wird, was zu einer anfänglichen Vergütung der Datenbanken führt, die proportional zu diesen Summen ist.

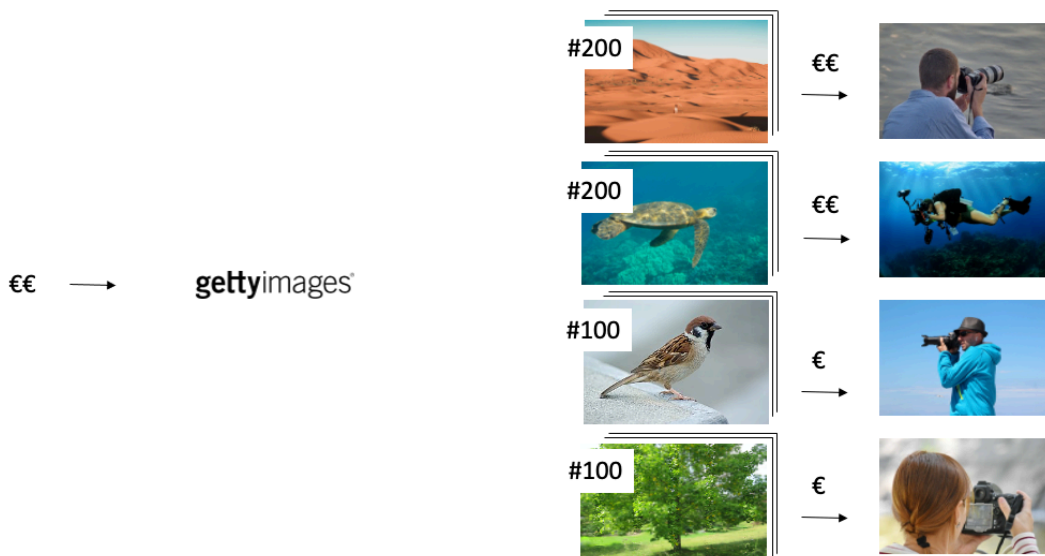


Messung der Ähnlichkeiten zwischen generierten Inhalten und Originalfotos zur Vergütung von Bilddatenbanken



Pay- to-Train“ -Methode unter den Autoren, und zwar nach dem Anteil ihrer einzelnen Werke, die zu den verkauften Datensätzen beitragen .

Anzahl der Fotos, um die Autoren zu bezahlen



Es werden weitere Verfeinerungen vorgenommen, wobei Bilder mit qualitativ hochwertigen Metadaten einen Bonus erhalten.

Brias Methode besteht daher aus zwei Schritten:

- Erstens dank eines Ähnlichkeitsalgorithmus zwischen Trainingsbildern und generierten Bildern
- Dann, dank eines Algorithmus des quantitativen Beitrags der Autoren zu den verkauften Datensätzen („ pay -to-train“)



Diese Hybridlösung ist:

- ökonomisch kompatibel mit dem Geschäftsmodell der generativen KI
- für zwei Jahre von Bilddatenbanken und Autoren akzeptiert
- wird derzeit auf andere Arten von Inhalten ausgeweitet: 3D, Musik, Sprache, animierte Bilder, Videos

Möglicherweise greift insbesondere das Startup ProRata auf diese hybride Methode zurück und bietet sie Medien an, die bereits Vereinbarungen mit KI-Betreibern haben. Dabei geht es darum, „die Ergebnisse der generativen KI, sei es Text, Bild, Musik oder Film, in ihre Bestandteile zu zerlegen, um herauszufinden, woher sie stammen, dann jedem Urheberrechtsinhaber einen Prozentsatz zuzuweisen und ihn zu bezahlen. entsprechend.“ Nähere Einzelheiten sind aus den öffentlichen Patenten des Mutterkonzerns nicht ersichtlich.

Diese Hybridmethode weist noch einige Mängel auf:

- Es ist nicht erwiesen, dass sich diese Lösung an alle Arten von Inhalten anpasst, beispielsweise an Spiele
- Es gibt keine Garantie, dass die Ähnlichkeit nicht zufällig ist. Mit anderen Worten: Ein Autor könnte bezahlt werden, obwohl seine Werke für die generierten Inhalte nicht nützlich sind und umgekehrt.
- Der Beitragsgrad eines Originalbildes wird ungefähr gemessen. Im obigen Beispiel erhält der Wüstenfotograf die gleiche Bezahlung wie der Meeresschildkrötenfotograf, da ihre Anzahl an Beiträgen zum Getty Images- Datensatz gleich ist, während die Wüstenfotos für den generierten Inhalt keinerlei Nutzen haben.

Es ist zu beachten, dass:

- selbst wenn die Verteilung der Vergütung für einen bestimmten generierten Inhalt nur annähernd erfolgt, verbessert sich ihre Gerechtigkeit wahrscheinlich mit der Menge des generierten Inhalts.
- durch Bria ist nicht erforderlich, sondern nur die Identifizierung der Trainingsdatensätze
- der Verteilungsalgorithmus selbst ist nicht öffentlich. Möglicherweise werden die Autoren künftig mehr Transparenz fordern, wahrscheinlich ist aber auch, dass im Gegenzug eine gewisse Geheimhaltung erforderlich sein wird, um opportunistischen Voreingenommenheiten vorzubeugen.
- Auch ist uns die durchschnittliche Vergütung der Autoren im Verhältnis zum Umsatz der generierten Inhalte, die unserer Meinung nach veröffentlicht werden könnten, nicht bekannt.

## Methoden der Datensatzähnlichkeit

Das Startup Sureel stellt Musikern ein Tool zur Verfügung, das den prozentualen Einfluss ihrer Arbeit auf einen KI-generierten Track misst.

Der Algorithmus, den er in seinem Patent von 1994 beschreibt, sagt die Vergütung eines Autors auf Grundlage des Einflusses aller seiner Werke voraus, gemessen beispielsweise anhand der Differenz zwischen den Gewichten des Basismodells und denen seiner Fein Anpassung durch diesen Satz. Ziel dieser Variante ist es, den Ähnlichkeitsfehler durch die Vergleichung einer größeren Anzahl von Werken abzumildern.

## Derzeit getestete Methoden

Über diese wenigen Variationen hinaus wetteifern die Forscher mit Einfallsreichtum, um neue Methoden zu testen:

Suche nach fairen Vergütungsmodellen

GESCHÄFT	VERFAHREN	FORSCHUNGSTEAM	DATUM	MOTIVATION
<p>OFFENE KI</p> <p>Princeton Universität</p> <p>Columbia Universität</p> <p>Harvard Universität</p> <p>Universität von Pennsylvania</p> <p>Universität von Kalifornien</p> <p>Berkeley</p>	Shapley	<p><b>Eine wirtschaftliche Lösung für die Urheberrechtsprobleme der generativen KI</b></p> <p>Jiachen T. Wang, Zhun Deng, Hiroaki Chiba-Okabe, Boaz Barak, Weijie J. Su</p>	<p>14. April 2024</p> <p>9. September 2024 (v4)</p>	<p>Shapley-Werte bieten einen prinzipiellen Ansatz zur Verteilung von Gewinnen auf der Grundlage des Nutzens.</p> <p>Das Dienstprogramm hilft dabei, das Ausmaß zu messen, in dem Datenquellen für die Generierung von Inhalten verantwortlich sind. Der Wert ist niedrig, wenn es unwahrscheinlich ist, dass das kontrafaktische Modell denselben Inhalt generiert wie das bereitgestellte Modell und umgekehrt.</p> <p>Der Nutzen kann als die Gesamtentschädigung interpretiert werden, die alle Mitglieder gemeinsam für die Bereitstellung ihrer Daten zum Trainieren des generativen KI-Modells verdienen.</p>
<p>Virginia Tech</p> <p>Massachusetts Institut für Technologie</p>	<p>In-Run-Daten Shapley</p>	<p><b>Data-Shapley in einem Trainingslauf</b></p> <p>Jiachen T. Wang, Prateek Mittal, Dawn Song und Ruoxi Jia</p>	<p>29. Juni 2024</p>	<p>Boaz Barak, ein Harvard-Professor für Sicherheit („Superalignment“), der auch bei OpenAI arbeitet, verliert Shapleys ideale Methode nicht aus den Augen, auch wenn dies Kompromisse bei der Genauigkeit bedeutet.</p> <p>Sein Forscherteam demonstriert im wirklichen Leben, wie der Shapley-Wert von Datensätzen wie Wikipedia oder Github für ein allgemeines KI-Modell berechnet wird, das mit einem GPT-ähnlichen Datensatz trainiert wurde.</p> <p>Diesem Team zufolge garantiert eine Ausrichtung der Vergütung an den quantifizierbaren Beiträgen jedes Datensatzes die größtmögliche Gerechtigkeit bei der Verteilung der Lizenzgebühren. Sie gibt an, dass sie die kostengünstigere In-Run-Data-Shapley-Variante bevorzugt.</p>
	<p>Trainingsmodellierung</p>	<p><b>TRAK: Modellverhalten im großen Maßstab zuordnen</b></p> <p>Sung Min Park, Kristian Georgiev, Andrew Ilyas, Guillaume Leclerc, Aleksander Madry MIT</p>	<p>3. April 2023</p>	<p>Aleksander Madry, Professor am MIT, ist bei OpenAI für AI Reasoning verantwortlich. Die Trak-Lösung basiert auf einer ungefähren Darstellung des KI-Modells, um dessen Entwicklung während des Trainings zu beschreiben und Vorhersagen mit den Trainingsdaten in Beziehung zu setzen, die sie beeinflusst haben.</p> <p>Diese Lösung hat sich bei Bildklassifizierern, die mit ImageNet trainiert wurden, bei Vision-Language-Modellen (CLIP) und linguistischen Modellen (BERT und mT5) als nützlich erwiesen.</p>

GESCHÄFT	VERFAHREN	FORSCHUNGSTEAM	DATUM	MOTIVATION
ADOBE Carnegie Mellon Universität Universität von Kalifornien, Berkeley	Datensatzähnlichkeiten	<b>Auswerten der Datenzuordnung für Text-zu-Bild-Modelle</b> Sheng-Yu Wang, Alexei Efros , Jun-Yan Zhu, Richard Zhan	8. August 2023	Da die Ähnlichkeitsmethode wahrscheinlich zufällige Ähnlichkeiten aufdeckt, kann diese Verzerrung abgeschwächt werden, indem die Methode auf zwei Bildsätze erweitert wird, einen trainierten und einen generierten. Die Forscher von Adobe berechnen daher den Beitrag des ersten Satzes, des Datensatzes , zu einem Satz synthetischer Inhalte
	Wasserzeichen	<b>ProMark : Proaktives Diffusions-Wasserzeichen für die kausale Zuordnung</b> Vishal Asnani , John Colomosse et al.	14. März 2024	Durch die Tätowierung wird die Kausalität zwischen Trainingsinhalten und synthetischen Inhalten verdeutlicht. Dadurch kann die rechtliche Zuordnung oder Zahlung an die Urheber zuverlässig erfolgen.
	Verlernen	<b>Datenzuordnung für Text-Bild-Modelle durch Verlernen synthetisierter Bilder</b> Sheng-Yu Wang, Aaron Hertzmann , Alexei A. Efros , Jun-Yan Zhu, Richard Zhang	Juni 2024	Bei dieser Methode wird das synthetisierte Bild verlernt, indem die vergessenen Bilder per Proxy trainiert werden, diejenigen Bilder mit erheblichen Verlustabweichungen nach dem Verlernvorgang identifiziert und als einflussreich gekennzeichnet werden.
HUAWEI Universität von British Columbia Simon Fraser Universität KTH Königliches Institut für Technologie	Föderierter Shapley	<b>Verbesserung der Fairness bei der Datenbewertung im horizontalen föderierten Lernen</b> Zhenan Fan, Huang Fang, Zirui Zhou, Jian Pei, Michael P. Friedlander, Changxin Liu, Yong Zhang	23. Mai 2022	Die Motivation der Dateneigentümer hängt teilweise von der Fairness der Zusammenarbeit und der Belohnung beim föderierten Lernen ab.

## Varianten und Weiterentwicklungen

Bei allen genannten Methoden gilt:

- Die Vergütungsparameter sind auf Grundlage von Arbitrierungen anhand der Nutzenindikatoren anpassbar: Größe, Datum, Ähnlichkeit, Anzahl der Metadaten, unabhängiger kommerzieller Erfolg usw.

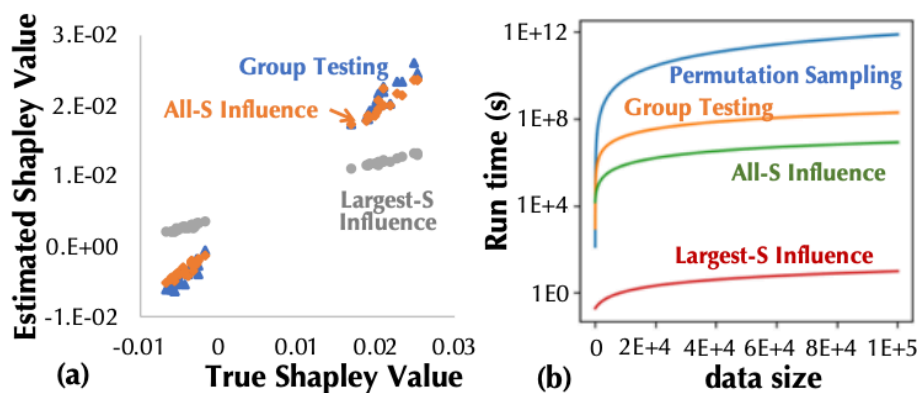
Indikatoren des Nutzens

FORSCHER, UNTERNEHMEN	MEDIEN	NUTZENINDIKATOREN
Piao , Chen et al.	Bild	Vektoren

FORSCHER, UNTERNEHMEN	MEDIEN	NUTZENINDIKATOREN
Wang, Deng et al.	Enzyklopädie, Wissenschaftliche Veröffentlichung, Bild, Code	Kontrafaktisches Modell (S) Vom $x(\text{gen})$ -Modell und dem kontrafaktischen Modell generierte Inhalte
Wang, Mittal et al.	Enzyklopädie, Akademische Veröffentlichung, Bild, Code, E-Mail, Forum	Halten des Validierungsdatenpunktsatzes $Z(\text{val})$
Deng, Zhang et al.	Musik	Durchschnittliche Geschwindigkeit der Ereignisse Durchschnittliche Höhe aller Events Summe der Zeitdifferenzen aller Ereignisse
Shutterstock + Bria	Bild, Vektorbild	Ähnlichkeitsbewertung Umsatz nach Datensatz Datenvolumen pro Datensatz und pro Autor Metadatenvolumen Rolle bei der Entwicklung der Originalmodelle Rolle durch Lizenzzahlungen im Zusammenhang mit zukünftigen generativen Lizenzierungsaktivitäten
Adobe Stock + Bria	Bild, Vektorbild	Ähnlichkeitsbewertung Bildzulässigkeit Bilddatum Anzahl der Downloads

- der Algorithmus ist hinsichtlich Präzision und Ausführungsgeschwindigkeit einstellbar. Der Shapley-Wert kann insbesondere berechnet werden
  - mit einer beliebigen Anzahl generierter Inhalte
  - mit beliebiger Granularität (nach Datensatz, Unterdatensatz, bis hin zu einzelnen Daten)
  - mit ähnlichen Funktionen

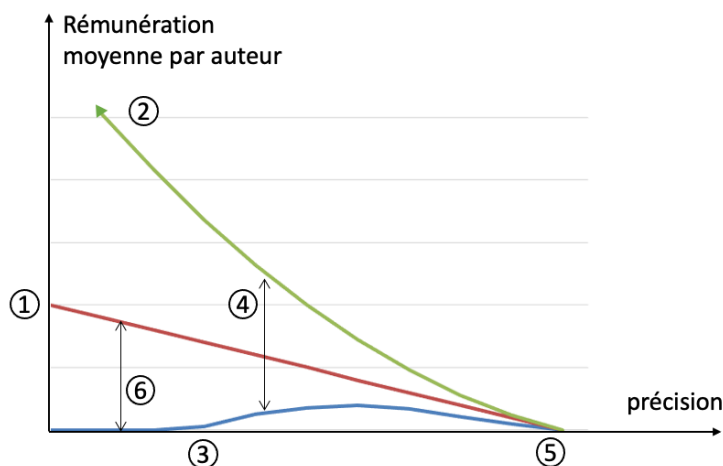
Drei Funktionen, die den Shapley-Wert approximieren



Beim derzeitigen Stand der Technik ist das Untersuchungsfeld breit genug, um durch die Nutzung Vergütungslösungen zu finden, die das Verhältnis zwischen Gerechtigkeit, Kosten und Präzision optimieren.

Auswirkungen der Präzisionskosten auf die Vergütung.

- (13) Bei der Genauigkeit Null entspricht die Vergütung pro Autor der Gesamtsammlung geteilt durch die Anzahl der Autoren.
- (14) Bei einer Genauigkeit von Null entspricht die maximale mögliche Vergütung eines Autors der Gesamtsammlung
- (15) Bei einer Genauigkeit von Null ist die minimale mögliche Vergütung eines Autors gleich Null
- (16) Unsicherheitslücke
- (17) Wenn die Kosten der Präzision die Gesamterfassung übersteigen, sind die Kompensations- und Unsicherheitslücke gleich Null
- (18) Bei gleichbleibender Erhebung sinkt die durchschnittliche Vergütung pro Autor mit der Genauigkeit



## 7 – Akzeptanz

**JDN**  
JOURNAL DU NET

**#FairlyShare contre le péché originel d’Internet** – septembre 2014, par Vincent Lorphelin, Gilles Babinet et al.

Ce qu’ils dénoncent ? Les plate-formes qui reposent sur l’industrialisation du travail gratuit, en dehors des règles de l’appel d’offre, du concours ou du bénévolat. Ils emboîtent ainsi le pas à [...] la levée de boucliers qui a suivi le rachat du journal de crowdsourcing Huffington Post par AOL pour 315 millions de dollars. [Lire](#)

Pay -to-Train“-Methode ist derzeit die am weitesten verbreitete Methode zur Vergütung von Autoren, deren Werke zum Trainieren allgemeiner KI-Modelle verwendet werden. Trotz teilweiser Transparenz und Fairness kam es bislang zu keinen Klagen oder Petitionen, wie man etwa bei der Aufklärung der „kostenlosen Arbeit“ oder dem Verkauf der Huffington Post (siehe Kasten) beobachten konnte.

Medienbanken nutzen „Pay -to-Train“-Modelle und nutzen ihre Datensätze

DATUM	MEDIE	DATENSATZANBIETER	Anbieter von KI-Modellen
25.10.22	Bild, Foto	Shutterstock	OpenAI
01/12/23	Bild, Foto	Shutterstock	Meta , Google , Amazon , Apple
12.03.23	Bild, Foto	Shutterstock	LG ( Exaone )
21.03.23	Bild, Foto	Getty Images	Nvidia
29.05.23	Avatar	Aufsteigende Kunst	
11.07.23	Bild, Foto	Shutterstock	OpenAI
08.08.23	Bild, Foto	Shutterstock	Nvidia (Picasso)
09/07/23	Bild, Foto	Getty Images Alamy Bria	

DATUM	MEDIE	DATENSATZANBIETER	Anbieter von KI-Modellen
13.09.23	Musik	AudioSparx	Stabilitäts-KI
13.09.23	Bild, Foto	Adobe Stock (Firefly)	
26.09.23	Bild, Foto	Getty Images (Getty Gen AI)	
10.04.23	Bild, Foto	Canva (Magisches Studio)	Offene KI (Dall-e) Google (Bild)
21.02.24	Bild, Foto	Superstock k Envato	
24.04.11	Bild, Foto	EyeEm Freiauswahl Fotoeimer	
27.06.24	Musik, Stimme	Universelle Musik* Sony Entertainment Warner Records	Google ( Youtube )
30.07.24	3D	Shutterstock	Nvidia (Erbauen)

\*Experimentieren

Die Kaufverträge für Datensätze der Medienbanken wurden von Reden begleitet, in denen für eine „respektvolle, verantwortungsvolle, sichere und faire“ Vergütung der Urheber geworben wurde.

Rede zur Gerechtigkeit bei der Ausbildungsvergütung

PLATTFORM	Bestand und Datensatz	ERKLÄRUNGEN
Nvidia	Getty, Shutterstock, Adobe	Nvidia : „Wir haben eine Plattform geschaffen, die es unseren Partnern ermöglicht, mit lizenzierten Daten zu trainieren, beispielsweise von Getty, Shutterstock oder Adobe“, sagte Huang. „Sie respektieren die Eigentümer der Inhalte. Die Trainingsdaten stammen aus dieser Quelle und alle daraus entstehenden wirtschaftlichen Vorteile können an die Ersteller zurückfließen. »
	Adobe	Adobe: Wir entwickeln generative KI verantwortungsvoll und mit den Entwicklern im Mittelpunkt . Adobe: Erstellen Sie mit der Gewissheit, dass Ihre Ergebnisse von höchster Qualität sind und mit Vorlagen erstellt werden, die für die sichere kommerzielle Nutzung konzipiert sind.

PLATTFORM	Bestand und Datensatz	ERKLÄRUNGEN
	Getty	Getty: Generative AI von iStock ist ein erschwingliches und kommerziell sicheres generatives KI-Tool [...] ohne die Angst, dass etwas rechtlich Geschütztes in den Datensatz gelangt und in ihrer Arbeit landen könnte
Meta	Shutterstock	Shutterstock : Fachwissen im Aufbau eines groß angelegten Ökosystems, das Mitwirkende mit Entwicklern belohnt und verbindet.
LG Exaone	Shutterstock	<p>Shutterstock : Das Eigentum an KI-generierten Inhalten kann nicht einer Einzelperson zugeschrieben werden. Stattdessen müssen die vielen Künstler entschädigt werden, die an der Erstellung jedes neuen Inhalts beteiligt waren. Unsere Kunden können KI-Bilder ohne Bedenken sicher und legal für den eigenen Bedarf lizenzieren. Darüber hinaus entschädigen wir Mitwirkende, deren Arbeiten unsere Modelle trainieren, sodass Sie unsere KI auch mit gutem Gewissen nutzen können. Unser KI-Bildgenerator bleibt die sichere, benutzerfreundliche und verantwortungsbewusst gestaltete Funktion, die Sie kennen und lieben</p> <p>LG: Das KI-Modell wird mit Millionen hochauflösender Bilder und Metadaten von Shutterstock trainiert und soll Textaufforderungen in Bilder umwandeln. Mitwirkende, deren Arbeiten zum Trainieren des Modells verwendet wurden, werden über den Contributor Fund von Shutterstock vergütet . Sie erhalten außerdem eine Vergütung, wenn unter Verwendung ihres geistigen Eigentums neuer generativer Inhalt erstellt und von Kunden lizenziert wird.</p>
Offene KI	Shutterstock	OpenAI : Die Daten, die wir von Shutterstock lizenziert haben, waren für das Training von DALL-E von entscheidender Bedeutung. Für OpenAI hat es höchste Priorität, dafür zu sorgen, dass die Creator Economy auch weiterhin lebendig bleibt . Schriftsteller, Künstler, Komponisten und andere Schöpfer haben im Laufe der gesamten Zivilisationsgeschichte unermesslich viel zur Gesellschaft beigetragen [...] OpenAI möchte die Schöpfer nicht ersetzen .
	Leinwand	Canva : Der Text-zu-Bild-Konverter von Magic Media, DALL-E von OpenAI und Imagen von Google Cloud [...] stellen sicher, dass ihre KI-Modelle fair und ethisch sind. „Wir teilen den Wert, den wir in Canva entwickeln , mit den Entwicklern, die all die Jahre bei uns waren. Mit dem neuen Creator Fund haben sie Zugriff auf einen Lizenzgebührenpool, der den KI-Kreationen entspricht. »
Google Cloud	Leinwand Shutterstock	Google Cloud: Shutterstock wurde auch zum führenden Innovator, indem es als erstes Unternehmen einen ethisch einwandfreien KI-Bildgenerator auf den Markt brachte, der jetzt mit Imagen auf Vertex AI erweitert wurde.
Google Youtube	UMG	Die Google-Tochter YouTube versucht, mit großen Musikgrößen Vereinbarungen zum Einsatz von KI zu treffen. Er möchte sicherstellen, dass „, Inhaber von Musikrechten für ihre Beiträge zu Trainingsdaten entschädigt werden.“ UMG und YouTube führen Gespräche über die Lizenzierung von Gesang und Melodien von Künstlern zum Trainieren von KI-Modellen.
Amazon SageMaker	Alamy Getty Envato	<p>Alamy : Im Mittelpunkt des Ethos des Fair -Diffusion- Programms steht die Verpflichtung, Fotografen, Künstler und Kreative gerecht zu entlohnen.</p> <p>Getty: Wir halten uns an die höchsten ethischen Standards und respektieren das geistige Eigentum und die Datenschutzrechte anderer.</p>
Sony		<p>Der Vorsitzende von Sony sagte: „Neue Produkte und Unternehmen, die auf der Grundlage von Musik entstehen, müssen mit der Zustimmung des Eigentümers und einer angemessenen Vergütung und Anerkennung entwickelt werden.“ Dabei gilt es zu verstehen, warum das Training von KI-Modellen durchgeführt wird, welche Produkte daraus entwickelt werden und welches Geschäftsmodell eine Monetarisierung der Nutzung der Werke des Künstlers ermöglicht.</p> <p>„Wenn ein generatives KI-Modell anhand von Musik trainiert wird, um neue musikalische Werke zu schaffen, die auf dem Musikmarkt konkurrieren, dann stellt das Training keine angemessene Verwendung dar. In diesem Fall kann die Ausbildung nicht ohne die Zustimmung, Nennung und Vergütung der Künstler und Rechteinhaber stattfinden. »</p>

## 8 – Patente

Bei der Erstellung dieses Berichts haben wir die folgenden Patente identifiziert:

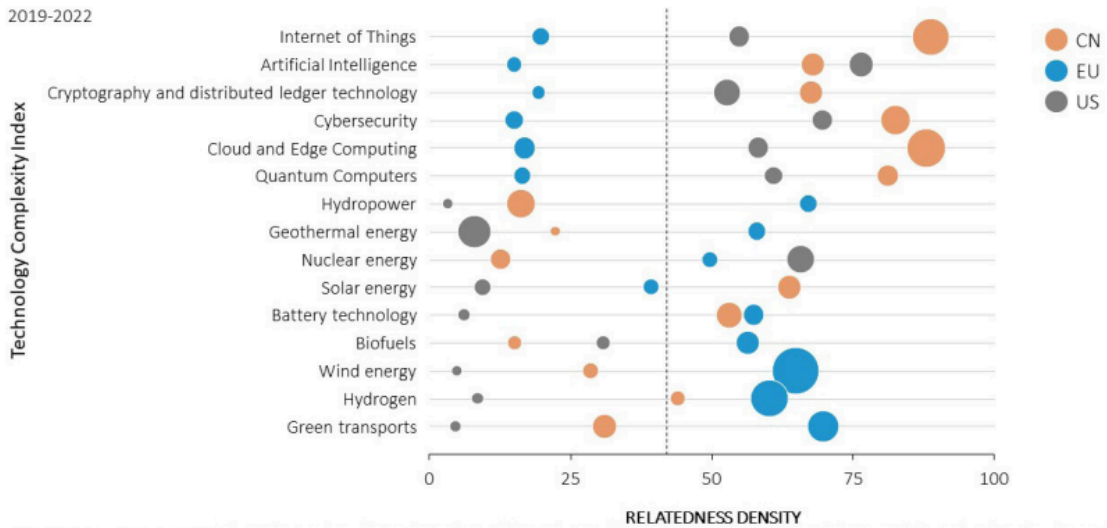
Patente im Zusammenhang mit der Vergütung von generativen KI- Trainingsarbeiten

TITEL	Referenz	LAND DES EINLEGERS	GELIEFERT	PIONIER
Systeme und Methoden kreativer Arbeit kollaborativer Systeme	US9100444	UNS	✓	✓
Bewertung des Stakeholder-Eigenkapitals in Gemeinschaftsprojekten	US10776889	UNS	✓	✓
Zuordnung generierter visueller Inhalte zu Trainingsbeispielen	US12033372	UNS	✓	
Modellbasierte Attribution für durch künstliche Intelligenz generierte Inhalte	US12013891	UNS	✓	
Generieren von Autorenvektoren	US10599770	UNS	✓	
Inputbasierte Attribution für durch künstliche Intelligenz generierte Inhalte	US2024419949	UNS		
Ausgabebasierte Attribution für durch künstliche Intelligenz generierte Inhalte	US2024419720	UNS		
System und Methode der Gebührenverteilung	US2023401550	UNS		
Erstellung, Nutzung und Zuordnung von Inhalten	WO2024166091	UNS		
Erleichterte Erstellung und Nutzung visueller Inhalte	WO2023084512	UNS		
Bundesweite lernbasierte Online-Kundenauswahlmethode und -system, Medium und Ausrüstung	Nr. CN118446287	China	✓	
Bildindexgenerierung basierend auf Ähnlichkeiten von Bildmerkmalen	Nr. CN103207879	China	✓	
Verfahren und System zur Bewertung des Beitragsgrads lokaler Bildmerkmale zu Gesamtmerkmalen	Nr. CN111652239	China	✓	
Methode und System zur Quantifizierung von Klientenbeiträgen im föderierten Lernen	WO2023036184	China		
Datenverarbeitungsmethode für die Verteilung mobiler Geräte auf Grundlage des Shapley-Werts des Bundeslernens	Nr. CN114912626	China		
Bundeslernanregungsmethode basierend auf Kooperation Sharpley-Wert	Nr. CN115965092	China		
Methoden und Systeme für föderiertes Lernen mit lokalen Prädiktoren	WO2024031564	China		
Methode und Kollaborationssystem	US 9723059	EU	✓	✓
Methode für einen erklärbaren Autoencoder und ein erklärbares generatives kontradiktorisches Netzwerk	US11948083	EU	✓	
Architektur für erklärbares bestärkendes Lernen	US11455576	EU	✓	
Mehrstufige Rückverfolgbarkeit von Beiträgen in einer Umgebung für kollektives Schaffen	US2015363743	EU		
Methode und Gerät zur Belohnung kollaborativer Kreation	EP2958068	EU		



Diese Liste zeigt, dass die Europäer, wenn es um die Vergütung der Rechteinhaber geht, nicht hinter China und den USA zurückliegen, während sie im Gegensatz dazu bei der künstlichen Intelligenz im Allgemeinen und der generativen KI im Besonderen eine Außenseiterposition einnehmen.

### The EU's position in complex (digital and green) technologies



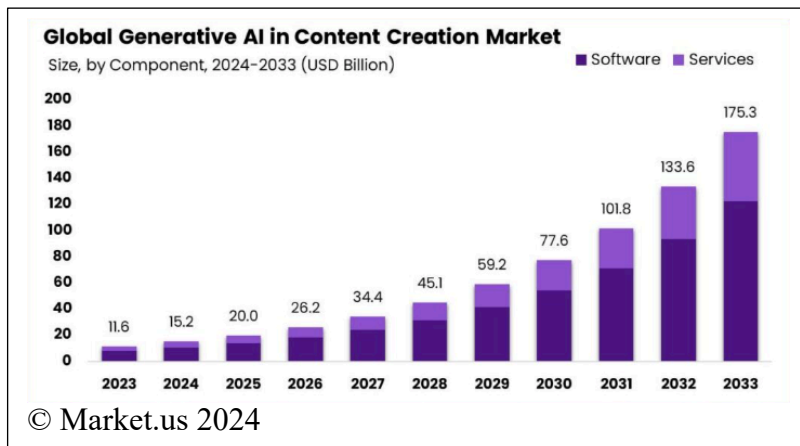
**Table 7 Top patent owners in GenAI modes (companies), 2014–2023**

	Images/ video	Text	Speech/ voice/ music	3D image models	Molecules/ genes/ proteins	Software/code	Other modes
Tencent Holdings (China)	607	565	551	102	57	41	464
Ping An Insurance (China)	262	600	599	26	33	24	223
Baidu (China)	395	465	441	81	26	18	166
IBM (US)	101	274	168	13	37	42	132
Alibaba Group (China)	142	213	144	36	2	10	133
Samsung Electr. (Republic of Korea)	173	140	226	34	10	5	57
Alphabet/Google (US)	138	107	200	25	15	15	72
ByteDance (China)	173	82	112	21	6	7	90
Microsoft (US)	78	194	151	18	6	22	42
BBK Electronics (China)	125	58	117	12	0	4	116
Netease (China)	78	76	80	37	0	11	102
NTT (Japan)	37	100	169	11	0	3	62
Huawei (China)	104	84	72	24	1	4	102
China Mobile (China)	79	67	83	10	0	7	89
State Grid (China)	75	54	31	6	1	2	144
Adobe (US)	190	79	34	30	0	9	20
Sony Group (Japan)	87	35	105	42	0	1	42
Siemens (Germany)	124	25	8	10	3	4	59
Ant Group (China)	45	82	29	3	0	6	59
Industrial and Commercial Bank of China (China)	40	62	49	5	0	16	48

Source: WIPO, based on patent data from EconSight/IFI Claims, April 2024.

**Perspektiven**

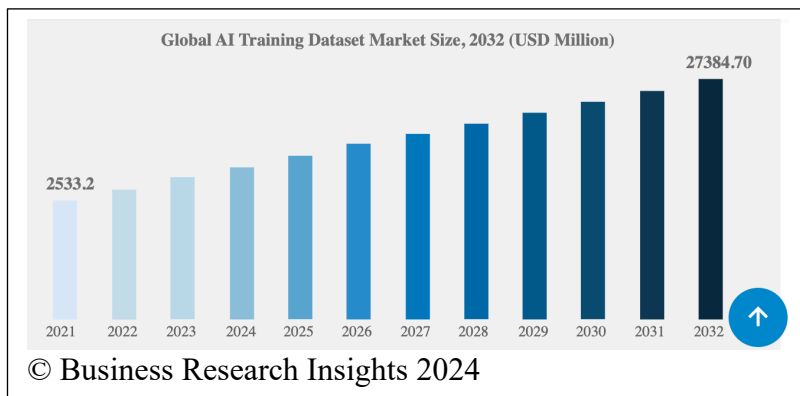
## 9 – Markt



Allein für das Kreativitätssegment wird das globale Marktvolumen für generative KI auf 11,6 Milliarden US-Dollar im Jahr 2023 geschätzt, im Jahr 2031 wird es voraussichtlich die Marke von 100 Milliarden US-Dollar überschreiten.

Um eine Vorstellung von der Größenordnung zu geben: Ein fairer Vergütungssatz von 15 % würde mehr bedeuten als die derzeitigen Einnahmen der OGCs (Organisationen zur kollektiven Rechteverwaltung) für künstlerische Rechte (Musik,

audiovisuelle Medien, Live-Auftritte, Literatur und bildende Kunst).



KI-Trainingsdatensätze spezialisierte Unternehmen organisieren die Sammlung, Annotation und Kuratierung der Daten sowie deren Anpassung an die Bedürfnisse von Lernalgorithmen.

Der globale Markt für KI-Trainingsdatensätze betrug im Jahr 2021 2,5 Milliarden US-Dollar und soll bis 2032 27 Milliarden US-Dollar erreichen.

### Vergütungsmodelle für Autoren generativer KI-Daten

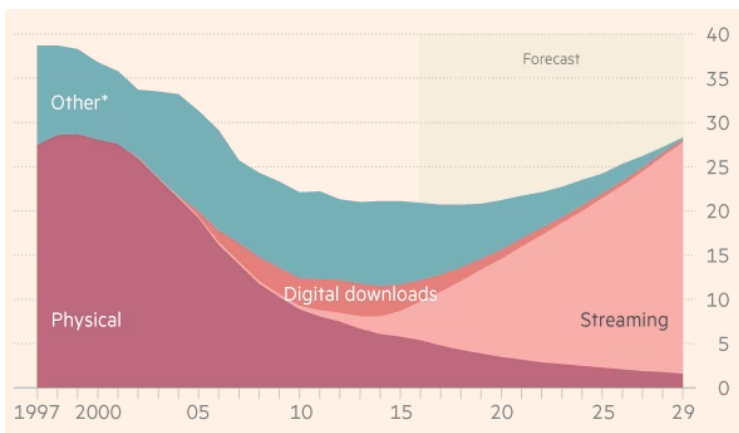
GESCHÄFTSMODELL	Unternehmen	AUTORENVERGÜTUNG
Schaben (1)	OpenAI , Midjourney , Stability , MistralAI	NEIN
Über der Daten (2)	Mturk , ClickWorker , Gigwalk , Prämisse	Zahlung bei Bestellung
Inhaltsbanken (3)	Shutterstock, Getty, Canva, Adobe	Bezahlte Ausbildung
Mediengruppen (4)	Axel Springer, Financial Times, Stack Overflow, Reddit, Le Monde	NEIN
Unternehmen Anzahl der Autoren (5)	die Autorengilde , Von Menschen erstellt, von Menschen geschaffene KI, fair trainiert, Dataset Providers Alliance	Faire Bezahlung

- Scraping : DALL·E wurde beispielsweise zunächst mit Hunderten Millionen beschrifteter Bilder aus dem Internet trainiert. Diese überwiegende Praxis ist Anlass für zahlreiche Klagen.
- Über der Daten: Hersteller von KI-Modellen haben begonnen, Risiken abzusichern und Datenversorgungsketten zu sichern, sowohl durch eine florierende Branche von Datenbrokern als auch durch Vereinbarungen mit Inhaltseigentümern, die entstanden sind, um die Nachfrage zu decken.  
Es entsteht eine Branche spezialisierter KI-Datenunternehmen, die Netzwerke aus kurzfristigen Vertragsarbeitern aufbauen, um auf Anfrage personalisierte Inhalte zu erstellen. Die Preise liegen zwischen 1 und 2 US-Dollar pro Einzelbild, zwischen 2 und 4 US-Dollar pro Kurzvideo und zwischen 100 und 300 US-Dollar pro Stunde für längere Filme. Der Marktpreis für Text beträgt 0,001 \$ pro Wort. Den Eigentümern der Fotos, Podcasts und medizinischen Daten werden etwa 20 bis 30 Prozent der gesamten Transaktionssumme ausgezahlt.
- Inhaltsbanken: Shutterstock , Getty, Canva oder Adobe verfügen über große Mediendatenbanken, die von Autoren bereitgestellt werden, die nach verkauften Datensätzen bezahlt werden .  
Shutterstock , der führende Anbieter von Bildern, hat Lizenzen für mehr als 100 Millionen US-Dollar oder 12 % seines Gesamtumsatzes an KI-Anbieter verkauft und wird voraussichtlich bis 2024 um 40 % wachsen. Diese Bildlizenzen, Videos und Musik werden einen adressierbaren Markt bilden von 10 Milliarden Dollar im Jahr 2030 und wird in diesem Zeitraum jährlich um 22 % wachsen.
- Mediengruppen: Datenbankbesitzer lizenzieren Datensätze , ohne zusätzliche Vergütung <sup>50</sup>Autoren
- Autorengesellschaften: Autoren möchten ihre Lizenzierungsoptionen für generative KI festlegen können. Ideen blühen auf wie die *der Authors' Guild* mit der Plattform „ *Created by Human* “ , wie *Human Native AI* , *Fairly Ausgebildete* , *Dataset Providers Alliance* oder die *KI-Kommission* in Frankreich. Zu diesem Zeitpunkt sind die Wirtschaftsmodelle noch nicht definiert und die Entwicklungsaussichten noch nicht quantifiziert.

Lizenzierungsvorschläge für Werke zur generativen KI

INITIATIVE	HAUPTVORSCHLÄGE
Die Autoren-gilde	<ul style="list-style-type: none"> <li>- Abschluss von Werknutzungsverträgen</li> <li>- Zahlung von Lizenzgebühren für die Nutzung geschützter Werke</li> <li>- Transparenz über die verwendeten Daten</li> </ul>
Von Menschen geschaffen	<ul style="list-style-type: none"> <li>- Von Menschen geschaffen “ -Label für Werke</li> <li>- Plattform, die es Autoren ermöglicht, Lizenzen zu autorisieren und zu kontrollieren</li> </ul>
Menschliche KI	<ul style="list-style-type: none"> <li>- Marktplatz für generative KI-Trainingsdatensätze</li> <li>- Tracking-Berechtigungen für jeden verwendeten Inhalt</li> </ul>
Gut ausgebildet	<ul style="list-style-type: none"> <li>- Label „Ziemlich Trainiert » für generative KI</li> <li>- Standardisierte Verträge für die Nutzung von Werken im KI-Bereich</li> </ul>
Allianz der Datensatzanbieter	<ul style="list-style-type: none"> <li>- Faire Vergütung</li> <li>- Standardisierte Verträge ( Pay -to-train, Umsatzsteuer, Abonnement, Sektorlizenzen)</li> <li>- Transparenz über die Verwendung von Trainingsdaten</li> <li>- Individuelle Berechtigungen pro Autor</li> </ul>
KI- Kommission	<ul style="list-style-type: none"> <li>- Öffentliche Plattform</li> <li>- Vermittlung der französischen Nationalbibliothek und des Nationalen Audiovisuellen Instituts</li> </ul>

Eine Parallele zum Musikmarkt ziehend, befinden wir uns derzeit in der „ Napster -IA“-Phase und warten auf Lösungen, die Spotify und Apple Music entsprechen . Ihre Geschäftsmodelle haben ab 2012 den Markt bereinigt, auf dem zuvor Streaming und Piraterie verwechselt wurden.



© Financial Times, Goldman Sachs

Ebenso sind wir davon überzeugt, dass der potenzielle Markt für die Aufwertung von Werken durch generative KI größer ist als der von OGCs und eine Chance für Rechteinhaber darstellt. Symmetrisch dazu könnten Betreiber generativer KI auf Datenbanken mit hochwertigen Inhalten zugreifen.

Dieses Potenzial, das derzeit alternativen Marktsegmenten zugutekommt, wird durch die

Entwicklung neuer Wirtschaftsmodelle vom Typ „Autoren-gesellschaften“ und ihre freiwillige Nutzung durch die Beteiligten freigesetzt.

Auch wenn diese Initiativen noch in den Kinderschuhen stecken, eröffnet die zunehmende Verbreitung eine neue Perspektive für „liquides Eigentum“.

## 10 – Auf dem Weg zu liquiden Vermögenswerten

### Interview de Vincent Lorphelin

Le Soir (Bruxelles), Janvier 2024



„Sam Altman und Elon Musk sagt, generative KI werde Arbeit überflüssig machen , also bereiten Sie sich auf ein bedingungsloses Grundeinkommen vor. Dies steht im Einklang mit der These vom Ende der Arbeit, die Jeremy Rifkin vor einigen Jahren formulierte . Dies steht auch im Einklang mit der Singularität , die nahelegt, dass innerhalb weniger Jahre die allgemeine KI gehen den Menschen in allen Bereichen übertreffen . Der Beweis ist, dass wir immer gesagt haben, dass KI niemals in die Berufe Einzug halten wird. kreativ (sie wird nie Autorin, Designerin usw. werden). Außer dass wir seit letztem Jahr dort sind. Die Tore der Bastion sind aufgebrochen . [...]

Umgekehrt ist die soziale Marktwirtschaft die DNA Europas, es ist die Wirtschaft , die Gesellschaft schafft . Arbeiten macht Sinn. Historisch und strategisch gesehen ist die

kollektive Rechtswahrnehmung in Europa doppelt so weit entwickelt wie in den USA . Kommen wir also zu dieser grundsätzlichen Frage des Interesses zurück. Allgemein , um die Singularitätsthese auszugleichen , das Interesse allgemein Europäische ist die Förderung des Modells dezentralisiert .

Stellen wir uns vor, dass dieses dezentrale Plattformmodell systemisch wird , zu einem System wird , so wie es heute bei zentralisierten Plattformen wie Gafam der Fall ist . In diesem Fall ist die Sache manipuliert werden , es handelt sich um geistige Rechte und nicht um Waren. Es handelt sich dabei um einen nicht marktwirtschaftlichen Wert . Der Ökonom Karl Polanyi hatte argumentiert , dass die Marktwirtschaft die Arbeit zur Ware gemacht habe , obwohl Arbeit nicht mit einer Ware gleichgesetzt werden kann.

Zentralisierte Plattformen werden soweit wie möglich versuchen, den geistigen Eigentumsrechten einen Marktwert zu geben. In einer dezentralen Logik werden Rechte in Prozenten und nicht in Währung ausgedrückt. Ökonomische Werte manipuliert werden, werden nicht mehr in Euro oder Dollar ausgedrückt. Es ist eine Wirtschaft , die nicht auf Geld, sondern auf Eigentum basiert . Um eine Parallele zu dem zu ziehen, was Emile Zola als "liquides Geld" bezeichnete (mit dem Übergang von einer Wirtschaft , in der der Reichtum nicht mehr aus dem Landbesitz, sondern aus dem Geldumlauf stammte), wird der Begriff des " liquiden Eigentums " vorgeschlagen, bei dem das Betriebskapital der Wirtschaft ist die Zirkulation von Rechten des geistigen Eigentums .

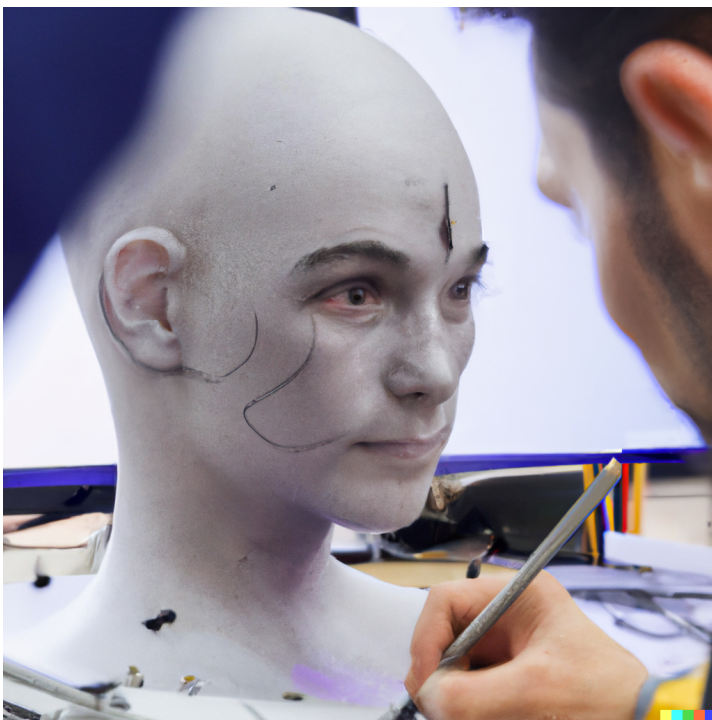
Heute haben wir alles so weit finanziert, dass ein Unternehmen seine Bilanzen täglich so präsentiert , als stünde es zum Verkauf , weil es alles in Geld misst. Allerdings gibt es immaterielle , nicht in der Bilanz ausgewiesene Vermögenswerte, die Werte ausdrücken, die nicht in Geld ausgedrückt werden können. Der Reichtum des Unternehmens geht über das hinaus, was man in Geld ausdrücken kann. Wenn wir die Eigenschaften einer Flüssigkeit manipulieren , müssen wir andere Indikatoren finden. Allmählich bewegen wir uns in Richtung einer Gesellschaft , die Reichtum einer anderen Größenordnung liquide macht .“ (Lesen Sie [hier mehr](#) ).

So wie das Plattformmodell von Uber das Konzept der Uberisierung inspirierte, nennen wir „Sacémisierung“ das Phänomen der Verbreitung des Wirtschaftsmodells von Urheberrechtsgesellschaften, das das Humankapital durch liquides Eigentum aufwerten wird.

## 11 – Fördern Sie die menschliche Intelligenz

Brad Smith, Präsident von Microsoft, begeht einen moralischen Fehler gegenüber dem Unternehmen. Über generative KI wie ChatGPT sagt er: „Nach dem Urheberrecht haben wir alle das Recht, zu lesen und zu lernen. Wir fragen uns nun, ob wir Maschinen die Möglichkeit geben können, auf die gleiche Weise zu lernen. Benehmen. Ich glaube, es gibt einen Imperativ gesellschaftlich, um dies zu ermöglichen.“

Dieser Eingriff mag trivial erscheinen. Wenn Maschinen das Internet saugen, um zu einem lächerlich niedrigen Preis Artikel und Illustrationen zu erstellen, warum sollten wir uns dann etwas vorenthalten?



Geben die New York Times und andere Rechteinhaber, die gegen diese Verwendung ihrer Inhalte Einspruch erheben, Microsoft nicht eine schlechte Chance? Wenn Maschinen kein Plagiat begehen, wo liegt dann der Schaden? Dies ist das inakzeptable Endergebnis dieser Rede.

### **Eine vorteilhafte rechtliche Unbestimmtheit**

die Gafam-Plattformen Unmengen kleiner Beiträge und fördern kostenlose Arbeit. Sie wechseln erneut die Dimension mit der KI, deren Gier die Eroberung neuer Gebiete erfordert, nämlich der Gebiete der Autoren.

Als Ausnahme hatte der Gesetzgeber den Begriff „Fair Use“ definiert, der es Google erlaubt, Auszüge privater Inhalte zu veröffentlichen. Da der Ausgang des Gerichtsverfahrens noch ungewiss ist, weiß der ehemalige Anwalt Brad Smith, dass die

öffentliche Meinung eine große Rolle spielen wird. Aus diesem Grund destilliert er das verfälschte geistige Bild vom „Recht der Maschinen zu lernen“. Sein „Imperativ“ gesellschaftlich ist nur eine vulgäre Anfechtung der Legitimität der Autoren, ihre Territorien zu verteidigen.

Er lenkt ab, indem er falsche ethische Fragen diskutiert. Ist die menschliche Intelligenz dazu verdammt, herabgestuft? Können wir KI Entscheidungen treffen lassen? Er könnte durchaus die Haltung eines weisen Mannes vor einem Elon einnehmen. Obwohl sie weniger höflich sind, sind sie beide Vorboten derselben Illusion: der, die menschliche Intelligenz zu übertreffen.



## **KI: eine Illusion der Überlegenheit ?**

Der beste Roman wird niemals die Lebenserfahrung ersetzen. Der beste „Videoanruf“ wird niemals ein reales Treffen ersetzen . Wie ausgefeilt das Modell auch sein mag, es entspricht nicht dem Original . Jeder von uns verfügt über eine Intelligenz der Welt, die immer die der Maschinen übertreffen wird , die über keine andere Intelligenz verfügen als die ihrer Vorbilder .

KIs erhellen unsere Intelligenz. Sie ermöglichen es der Umweltpolizei, Umweltverschmutzer zu identifizieren , dem Lehrer, die Schwierigkeiten von tausend Schülern zu verstehen oder dem Labor, die Auswirkungen eines Medikaments vorherzusehen .

die Illusion einer unmittelbar bevorstehenden Überlegenheit der Maschinen zu verbreiten . Es ist eine Leugnung der streng menschlichen Arbeit, die ihre Inspiration gerade aus der Welt und ihren Modellen bezieht , wenn der Künstler etwas Unausgesprochenes enthüllt , wenn der Fahrer schlecht reagierende Hebel kompensiert , wenn der Arzt eine zu offensichtliche Diagnose anzweifelt oder wenn der Richter hat eine vertrauliche Überzeugung.

## **Die Verteidigung der Autoren**

Die Illusion eines Wettbewerbs zwischen Intelligenzen zu nähren und Maschinen Menschenrechte zuzuschreiben, ist ein Fehler der Gesellschaft . Im Gegenteil, wir müssen die Vorrangstellung der menschlichen Intelligenz bekräftigen und die Integrität ihrer Territorien garantieren.

Brad Smiths Illusion wäre bedeutungslos, wenn es die weltweit führende Persönlichkeit auf dem Gebiet der KI nicht gäbe . Von dieser Illusion überzeugt , behauptet die Interministerielle Kommission für KI, dass die Maschinen den Menschen übertreffen werden . Sie verteidigt glücklicherweise die gerechte Entlohnung der menschlichen Schöpfung , ohne jedoch zu erkennen, dass es sich dabei um die Wertschätzung handelt sogar von dieser Genauigkeit, die manipuliert wird .

Europa hat mit dem AI ACT bereits beschlossen , das Urheberrecht zu schützen . Sie muss nun den Kataster der von den KIs ausgebeuteten Gebiete erstellen und deren Pacht verteidigen . ( [Lesen Sie](#) mehr in Les Echos vom 6. Mai 2024)



## 12 – Eine sacemisierende Lösung für Europa

### Entretien avec Jean-Paul Betbèze

Economiste, Membre du Comité scientifique de la Fondation Robert Schuman



### Was sollten die Kernpunkte der „Sacemisierung“ sein?

sacemisierende ‘ Lösung für Europa sollte:

- Basieren auf dem geistigen Eigentum der Autoren, wie die OGCs. Das geistige Eigentum der Autoren muss bewahrt und geschützt werden, da es im Zuge der gegenwärtigen Kommunikationsrevolution andernfalls verloren gehen oder bleiben könnte. Die Kommunikation wird globaler und bildet die Grundlage für den tiefgreifenden Wandel, den wir derzeit mit der Vervielfältigung von Nachrichten in alle Richtungen erleben. Der Beitrag eines jeden muss anerkannt werden, es sei denn, wir trocken die gegenwärtige globale Bewegung durch eine übermäßige Konzentration zugunsten bestimmter Autoren aus, während die anderen in ihrer Arbeit und ihren Ideen keinen Vorteil, keine Anerkennung, keine Vergütung finden. Aufgrund der Lücken in der

Kapitalverfügbarkeit, von denen vor allem die USA profitieren, besteht die Gefahr einer Monopolisierung der offensichtlichen Schöpfungsquellen.

- Erheben Sie Lizenzgebühren für die Nutzung von Werken durch KI. Die „Sacemisierung“ ermöglicht uns nicht nur die Rückverfolgung, sondern auch die Identifizierung und Einschätzung von Quellen, selbst wenn diese minimal erscheinen. Durch dieses Verfahren bleiben alle Ursprünge erhalten und damit auch alle Neuerungen und Verbesserungen erkennbar: nichts geht verloren, alles bleibt erhalten.
- Erkennen Sie Kreativität und unterscheiden Sie sie von Amateurinhalten. Bei der Verbreitung der Werke ist es wichtig, diejenigen hervorzuheben, die originell sind, sie in den Vordergrund zu rücken und zu fördern.
- Implementieren Sie individuelle Anreizmechanismen, damit der Preis für Kreativität auf dem Markt wettbewerbsfähig und mit den Produktionskosten vereinbar ist. Es besteht die Gefahr einer Vervielfältigung der Werke, insbesondere da ihr physischer Produktionsaufwand offensichtlich gering ist. Eine Abrechnung, die sich ausschließlich an den Produktionskosten orientiert, birgt die Gefahr, entmutigend zu wirken. Der große Vorteil der „Sacemisierung“ besteht darin, dass man sich von einer Preislogik, die an die Produktionskosten gekoppelt ist, abwendet und sich einer anderen Logik zuwendet, nämlich der Logik der Märkte, die auf der Bedeutung des Erfolgs und der Wiederholung der Botschaft basiert.

## **Was ist der Vorteil eines Marktes, der von „ Sacemisierung “ getrieben wird?**

In dieser Marktlogik ist es die gemessene Nachfrage, aus der sich die erwartete Nachfrage ergibt, die die neuen Preise und damit die Margen schafft. In diesem Zusammenhang können die Margen und damit die Grenzgewinne und damit die Bewertungen nur steigen. Dieser Unterschied zwischen den niedrigen Kosten, die von Suchmaschinen erkannt werden, und den Bewertungen, die somit eine objektive Grundlage haben und durch Tools vom Typ Sacem aktualisiert werden, ist die Grundlage der vorgeschlagenen Logik. Dadurch können die Beiträge aller hervorgehoben werden, insbesondere im europäischen Kontext.

## **Und für kreative Tätigkeiten?**

Wir dürfen nicht vergessen, dass es immer einen monopolistischen Wettbewerb gibt: Er strebt eine Verringerung der Zahl der Wettbewerber an, bis nur noch einer übrig bleibt: das Monopol. Dort gilt mehr als anderswo, dass der Gewinner den gesamten Einsatz erhält. Dem muss durch die „ Sacemisierung “ begegnet werden, indem eine bessere Verteilung der Vergütungen ermöglicht wird, bevor ein einzelner Monopolist auftritt. Dies kommt der Forschung, der Innovation und der Schöpfung offensichtlich zugute, denn wir wissen, dass sich die Monopolsituation gegen den Monopolisten selbst wendet, wenn die Marktsituation auf lange Sicht austrocknet, weil ihre Dynamik zu sinkenden Gewinnen führt. Die „ Sacemisierung “ respektiert somit die Schöpfungen und verlangsamt die gnadenlose und letztlich zerstörerische Selektion des Monopols selbst.

Auf dem Markt der Ideen ist Geschwindigkeit tödlich, entscheidend ist die Erinnerung an Beiträge, verbunden mit ihrer gerechten Vergütung. Dies ist heute möglich und verständlicherweise notwendiger denn je. Es geht um die Rettung der Schöpfung durch den Fortschritt, den die Technologie bietet. »

Michelle Bergadaà, Professeure émérite à l'université de Genève.

Co-autrice de *Réinventer l'intégrité académique à l'ère de l'intelligence artificielle* (à paraître, EMS, février 2025)



### **Sie sagen, das KI -Gesetz betrifft nicht nur Urheber, sondern auch Forschende, warum?**

MB: „Wir müssen darüber nachdenken, das KI -Gesetz anzuwenden , um Autoren „authentischer Forschung“ zu entschädigen.“ Frankreich bezahlt seine Forscher schlecht, was deren Abwanderung ins Ausland oder in die Privatwirtschaft verlangsamen würde. Es kostet die Gesellschaft viel mehr, ihre jungen Forscher mit bürokratischen Hürden zu überziehen, als sie gut zu bezahlen. Viele verlieren die Motivation und leisten zusätzliche Lehrstunden, statt zu forschen. Wenn der Staat ihnen also keine angemessene Bezahlung bieten kann, muss dies über das System der Produktion von Artikeln, Büchern und Berichten erfolgen. Dies muss der Eckpfeiler der europäischen Strategie für generative KI sein. Mit dem zusätzlichen Bonus eines automatisierten Betrugserkennungssystems.

### **Hat es Betrug nicht schon immer gegeben?**

Die Opioidkrise in den USA [allein Fentanyl und verwandte verschreibungspflichtige Medikamente waren im Jahr 2021 für schätzungsweise 71.000 Todesfälle durch Überdosierung verantwortlich] entstand aufgrund eines einzigen, falschen, falsch zitierten und nachgedruckten Artikels. Das war lange vor der künstlichen Intelligenz. Die Verbreitung betrügerischer Artikel erfolgt bei KI vor allem über

neue Artikel. Dies ist ein neues Modell der Verbreitung des Falschen.

Heute existieren drei Zeitlichkeiten nebeneinander. Erstens ermöglicht die unmittelbare und emotionale Zeit (von Trump und der Generation Z) nicht nur die Produktion falscher Informationen, sondern auch falscher Autoren, die ebenso glaubwürdig sind. Zweitens erhält die ewige Gegenwart der Gemeinschaften von gestern eine alternative Welt aufrecht, in der die Erde flach ist und Bleichmittel Covid eliminiert . Die dritte lineare Zeit wird von Elons materialistischer Technologie angetrieben Moschus . KI ist ein Betrugsbeschleuniger, der in den Zwischenräumen dieser Zeitlichkeiten gedeiht.

### **Wie soll man vorgehen?**

Pubpeer und die Retraction Watch Database versuchen, einzelne Täter einzeln zu fassen. Doch dauerte es drei Jahre, bis die Untersuchung abgeschlossen war und die Zeitungen einen Widerruf erwirkten.

Dieses Phänomen muss systematisch mit Integritätsradaren, einer Anordnung für Betrüger, das erhaltene Geld zurückzugeben, und Ausschluss KI-Datenbanken. Die Zukunft unserer Gesellschaft sind die jungen Menschen, die wir ausbilden. Helfen wir ihnen, indem wir eine klare Haltung einnehmen und ihre originellen Kreationen durch generative KI vergüten. Dies würde viele junge, ehrliche Forscher motivieren, Wissen zu schaffen.

# Empfehlungen

## 13 – Die Controv3rse-Lösung

Sam Altman, Gründer von Open AI, möchte „neue Geschäftsmodelle schaffen, die es Entwicklern ermöglichen, Einnahmen zu erzielen.“ »

Wir haben gesehen, dass das Uber-Datenmodell entsteht, um die Degeneration der generativen KI zu verhindern, dass dies jedoch eine Bedrohung für das Humankapital darstellt. Das Wirtschaftsmodell muss sich auf OGCs stützen, um eine Perspektive der Neubewertung dieses Kapitals durch liquide Mittel und Sacemisierung zu eröffnen .

Das Geschäftsmodell, das wir Sam Altman und anderen Betreibern generativer KI empfehlen, ist folgendes:

### Controv3rse-Lösung

EMPFEHLUNG	OFFENE KI	OGC
1 - Reservieren	15 % des generativen KI-Umsatzes für eine faire Vergütung von Autoren und Rechteinhabern reservieren	Bieten Sie Autoren eine Plattform, um gegebenenfalls der Nutzung ihrer Werke durch generative KI zu widersprechen oder für jedes Werk eine Lizenz auszuwählen.
2 – Segmentierung	Identifizieren Sie die wichtigsten OGC nach Marktsegment (Medium/ Datensatz * Land)	Beanspruchen Sie ein Marktsegment (Medium/ Datensatz * Land)
3 - Methode	Veröffentlichen Sie die Zuordnungsmethode des OGC und erläutern Sie die Begriffsklärung im Falle einer Wiederherstellung.	Aushandeln von Koordinierungs- oder Gegenseitigkeitsvereinbarungen zwischen OGCs
4 – Keine Namensnennung	Identifizieren Sie einen Fonds nach Land für nicht zugewiesene Beträge	Forderungsmanagement des Fonds
5 - Berichte	Veröffentlichen Sie die Zusammenfassung in ausreichendem Detail  Senden Sie jedem identifizierten OGC einen monatlichen Bericht mit den Kennungen der Autoren, den Einzelheiten zur Verwendung ihrer Werke nach synthetischem Inhalt und dem zugewiesenen Wert	Senden Sie dem Betreiber die Liste der Werke, die von der generativen KI nicht verwendet werden können
6 - Zahlung	Verteilen Sie die Reserve an OGCs und Fonds	Verteilen Sie die eingenommenen Beträge an Autoren und Rechteinhaber

# 14 – Beginnen Sie die Verhandlung

## Vorgehensweise

Um Vereinbarungen zwischen Betreibern generativer KI und CMOs zu treffen, empfehlen wir den Beteiligten, die folgenden Schritte zu dokumentieren:

### Schritte der Controv3rse-Lösung

EMPFEHLUNG	VERFAHREN	BEST PRACTICES, VERGLEICHBAR
1 - Prototyp	Diskutieren Sie über eine zu perfektionierende technische Lösung, um den Gesprächspartnern die Umsetzung eines Ergebnisses zu erleichtern und die Komplexität der Debatte zu reduzieren (siehe unseren Demonstrator).	Demonstrator, Versuch und Irrtum Methoden
2 - Sicht	Weitsicht entwickeln: Dezentralisierung vs. zentralisierte Plattformen, Sacemisierung vs. Uberisierung der KI, Arbeit der Autoren vs. Kommerzialisierung, Absurdität des Maschinenrechts, liquides Eigentum, Nutzenmetriken	Europäischer Rahmen: Soziale Marktwirtschaft, Ökonomie der Kreativität, Erbe und kulturelle Vielfalt
3 - Faire Vergütung	Legen Sie eine angemessene Vergütungsrate fest, die sich an der Nützlichkeit der Arbeiten aller Autoren für alle generierten Inhalte orientiert. Wir bieten einen Startwert von 15 %.  Schlagen Sie Indikatoren für die Nützlichkeit von Beiträgen für die Zuordnung einzelner Autoren sowie eine Methode zur Verfeinerung dieser Zuordnung im Hinblick auf künftige technische Fortschritte vor.	Spre Richtlinie 2006/115/EG (13) Leistungsschutzrechte (73)  Yacast- Statistiken für Spré
4 – Verhaltenskodex des AI-Büros	Hochmodernes Observatorium <ul style="list-style-type: none"> <li>- Methoden: Shapley-Wert, Pay -to-train, Ähnlichkeit</li> <li>- Schiedsverfahren: Parameter, Granularität, Nutzenindikatoren</li> <li>- Options- und Lizenzierungsplattformen</li> <li>- Identifizierung der Rechteinhaber und Autoren</li> <li>- Verarbeitung nicht zugewiesener Beträge</li> </ul> <p>Ausreichend ausführliche Zusammenfassung:</p> <ul style="list-style-type: none"> <li>- Inhalt</li> <li>- Granularität</li> <li>- Format</li> </ul> <p>Meldungen an Rechteinhaber:</p> <ul style="list-style-type: none"> <li>- Datensatzansprüche von Rechteinhabern</li> <li>- Zuteilungs- und Verteilungsmethoden durch Rechteinhaber</li> <li>- Bearbeitung von Rückforderungen</li> <li>- Daten</li> </ul> <p>Geheimer Inhalt: Informationen, die in Zusammenfassungen und Berichten nicht offengelegt werden</p> <p>Progressive Normalisierung</p>	Verhaltenskodex gegen Desinformation  Empfehlungen der Autorengilde  Mission der CSPLA          ISWC, ISAN
5 – Patente	Sicherung der französischen und europäischen Position durch ein Portfolio dedizierter Patente in öffentlich-privater Zusammenarbeit	INPI Alliance PI-Programm



## Faire Bezahlung

Eine gerechte Vergütung durch generative KI könnte sich an dem Konzept des Spré orientieren . In Frankreich erhebt diese Organisation Lizenzgebühren auf der Grundlage des Umsatzes der Nachtclubs zugunsten der Musiker. Eine repräsentative Auswahl von 120 von ihnen ist mit einer von Sacem und Spré finanzierten „Yacast“-Box ausgestattet . Yacast sammelt Wiedergabelisten mit gestreamter Musik und veröffentlicht Statistiken nach Künstler, Label, Genre und Anzahl der Sendewochen. Nach dem gesetzlichen Abzug von 9 % verteilt der Spré die Abgabe an die OGCs (SCPA 50 %, ADAMI 25 % und SPEDIDAM 25 %) und teilt ihnen die Yacast- Statistiken mit , um die Endbegünstigten (darstellende Künstler, Produzenten, Aktionen von allgemeinem Interesse) zu vergüten. .

## Fairer Preis

In seinem Bericht „*Der Autor und der Akt der Schöpfung*“ aus dem Jahr 2020 beklagte Bruno Racine die Tatsache, dass „die anteilige Vergütung der Autoren nur in Ausnahmefällen 10 % der Betriebseinnahmen erreicht oder übersteigt“. Er führte diese Beobachtung auf „die Beziehung zurück, die den Künstler-Autor mit den nachgelagerten Akteuren (Verleger, Händler, Produzenten) verbindet [die] zutiefst unausgewogen erscheint, was ihn in vielen Fällen dazu bringt, die Idee von Vertragsfreiheit.“ Um dem abzuhelfen, wollte er „die Künstler-Autoren kollektiv stärken, durch die rasche Organisation von Berufswahlen, die es ermöglichen werden, dem Nationalen Rat der Künstler-Autoren Substanz und Legitimität zu verleihen, der als Rahmen dienen soll für Kollektivverhandlungen mit den Rundfunkveranstaltern" sollen durchgeführt werden, um "ein vertragliches System zur Vergütung der Arbeitszeit des Autors zu etablieren, wobei der Sponsor Vorrang hat".

Da die Idee einer Neubewertung des Arbeitswerts im Fall der generativen KI nicht anwendbar ist, schlagen wir vor, den fairen Satz als den Nutzenwert aller Beiträge der Autoren im Verhältnis zur Geschäftszahl der generierten Inhalte zu definieren. Auf diesem Prinzip basiert der Shapley-Wert, und der Ökonom Ernst Fehr schätzt, dass der Wert, den die Presse der Google- Suche bringt, 14 % beträgt . Seit der berühmten Frage „War dieser Kommentar für Sie hilfreich ? » Mit der von Amazon eingeführten Strategie verbessern alle kommerziellen Websites die Qualität ihres Angebots durch die Vervielfachung der heute allgemein gebräuchlichen Nutzenkennzahlen.

Die Festlegung der Steuerbemessungsgrundlage auf der Grundlage des Umsatzes der generierten Inhalte würde es ermöglichen, dass die gerechte Vergütung des Autors nicht von der mehr oder weniger starken vertikalen Segmentierung des Sektors abhängt, zum Beispiel: Autor / Plattform / Datensatz / allgemeines Modell / Feinabstimmung / Schnittstelle / Generation.

## 15 – Lizenz – Danksagungen

Diese Notiz und ihre Tabellen sind unter der CC-BY-Lizenz zugänglich, vorbehaltlich der gerechten Aufteilung der Früchte ihrer Nutzung durch generative KI, gewährt durch den Verein Controv3rse, RNA W751267044.

Dieses Dokument wurde automatisch von Google Translate übersetzt. Im Falle von Unklarheiten ist die französische Version maßgebend.

Die Inspiration stammt aus den Überlegungen und Diskussionen zwischen Mitgliedern der Ministerien für das KI- Gesetz , denen der unabhängigen Denkfabriken Controv3rse und Institut de l' Iconomie sowie

anderen Persönlichkeiten, die für ihre Expertise anerkannt sind. Sie erweitert die Studie, die zuvor für die parlamentarische Mission Pradal -Rambaud durchgeführt wurde, sowie die in *Le Monde* und *Les Echos* veröffentlichten Kolumnen . Vincent Lorphelin, Gründer und Co-Präsident dieser Think Tanks und Autor dieser Notiz, dankt ihnen herzlich für ihre Ideen und die Qualität dieser Diskussionen.

Präsidenten und Berichterstatter der Parlaments- und Ministermissionen:

Alexandra Bensamoun , Professorin für Rechtswissenschaften an der Universität Paris Saclay, ehemaliges Mitglied der Kommission für künstliche Intelligenz

Joëlle Farchy , Universitätsprofessorin, Direktorin von m2ecn und PcEn- Lehrstuhlinhaberin Universität Paris 1 Panthéon-Sorbonne, Mitglied der CSPLA

Philippe Pradal , Abgeordneter für den dritten Wahlkreis der Alpes Maritimes, Nationalversammlung

Stéphane Rambaud, Abgeordneter für den dritten Wahlkreis Var, Nationalversammlung

Lionel Ferreira, Master of Requests beim Staatsrat

Julie Groffe -Charrier, HDR-Dozentin für Privatrecht an der Jean-Monnet-Fakultät

Bastien Blain, Juniorprofessor für Wirtschaftswissenschaften, Daten- und Verhaltenswissenschaftler, Universität Paris 1 Panthéon-Sorbonne

Experten und Unternehmer:

William Bailey  
Mitbegründer und Präsident, Bolero Music

Emmanuel Benazera  
Vorsitzender und CEO, Jolibrain

Jean-Paul Betbèze  
Ökonom , Betbèze Conseil

Michelle Bergadaà  
Präsident des Instituts für Forschung und Maßnahmen gegen akademischen Betrug und Plagiat

Michel Bokobza  
Missionsmanager, Collège de Paris

Laurent Cervoni  
Direktor für Forschung und Innovation, Talan

Charles-Eric de la Chapelle  
Gründer, Myriad Data

Jean-Philippe Clair  
Direktor für Marketing, Kommunikation und Innovation, Keyrus

Alice Coatalem  
Mitbegründer von CogNeed und außerordentlicher Professor an der Universität Paris-Dauphine PSL

Christophe Collet  
Gründer, AskLocala

Frederic Dayan  
Gründer, Exactcure

Guillaume Desveaux  
Mitbegründer von Aleia und Administrator der AI Cargo Foundation

Jonathan Dory  
Gründer , Live Crew



William Eldin  
Gründer, XXII

Stephanie Flacher  
Mitgründer, Logion Network

Raphael Frisch  
Mitgründer, HawAI.tech

Alain Garnier  
CEO, Jamespot

Anthony Graveline  
Gründer, Disaitek

Philippe Guguen  
Präsident, Sorga - Karte Emulsion

Francis Hachem  
Gründer, Codenekt

Rodolphe Hasselvander  
Gründer, Blue Frog Robotik

Brice Hoarau  
Gründer, Semdee

Matthias Houllier  
Mitgründer, Wintics

Julia B.  
Rechtsanwalt, IPso

Jean -Luc Latz  
Gründer, Oktal -SE

Frédéric Lefebvre- Naré  
KI-Datendirektor, Niji

Youness Lemrabet  
Gründer, Everysens

Jacques Levy- Vehel  
Präsident von Case Law Analytics und  
Forschungsdirektor, INRIA

Olivier Laborde  
Leiter für Innovation und digitale  
Transformation, BPCE

Alexandre Leforestier  
Gründer, Panodyssey

Sixtine Lorphelin  
KI-Ingenieur, UTC, Student am INSEAD

Aymeric Masurelle  
Mitbegründer , Spoon.ai

Pierre Miralles  
Mitbegründer, Footovision

Emmanuel Moyrand  
Mitgründer, France Meta

Clement Merville  
Gründer, Manzalab - Teemew

Edward von Miollis  
Gründer, Polycube

Nathalie Nevejans  
Professor für Privatrecht und künstliche  
Intelligenz, Universität Artois

Remy Özcan  
Mitgründer, Crypto4All

Jean-Jacques Quisquater  
Professor für Kryptographie,  
Polytechnique Louvain, Kolumnist für Le  
Monde

Jean-Michel Salomon  
Präsident, Society of Game Authors

Frederic Soufflet  
Mitbegründer, Haapie

Clement Tequi  
Mitbegründer, Terno

François-Xavier Thoorens  
CEO, Vaultys

Arnaud Touati  
Rechtsanwaltsanwältin, #Hashtag


Christophe Tricot  
Mitbegründer, La Forge

Killian Vermersch  
Mitgründer und CEO, Golem.ai

# La soluzione Controv3rse

Un modello economico per migliorare il capitale culturale e umano attraverso  
l'intelligenza artificiale generativa



 Adobe Firefly

C O N T R O V **E** R S E

**Riepilogo**

Controv3rse è un think tank indipendente, formato da 70 imprenditori ed esperti di intelligenza artificiale generativa. Questo rapporto offre una soluzione in risposta all'affermazione di Sam Altman, fondatore di OpenAI, che sostiene un nuovo modello economico che premi i creatori di fronte all'ascesa dell'intelligenza artificiale generativa.

Va detto che il conflitto tra gli autori, le cui opere vengono saccheggiate, e gli sfruttatori dell'intelligenza artificiale generativa, che hanno bisogno di chiarezza e stabilità giuridica, non verrà risolto in tribunale prima di almeno 10 anni.

Nel frattempo, la proliferazione dell'intelligenza artificiale generativa diluisce l'originalità e cancella le tracce delle opere originali. Allo stesso tempo, le piattaforme di micro-tasking rispondono alla massiccia domanda di dati a scapito dei modelli basati sulla creatività e sul copyright, come SACEM. Il capitale culturale e umano, un tempo centrale nella strategia europea, è doppiamente minacciato dall'estrazione.

Tuttavia, l'economista premio Nobel Jean Tirole sottolinea che la ricchezza nazionale si basa sulla promozione dell'innovazione, sostenuta dalla proprietà intellettuale come leva per la creatività. Per preservare il capitale culturale e umano stimolando al contempo l'innovazione, il nuovo modello economico deve reinventare quello delle società degli autori.

Nel 1953 Lloyd Shapley introdusse un concetto chiave per distribuire equamente le vincite in un gioco cooperativo, basato sul contributo di ciascun giocatore. Il metodo derivato, ideale per remunerare gli autori dell'IA generativa, ha purtroppo un costo computazionale proibitivo. Esistono alternative, come il modello “pay-to-train” utilizzato dalle banche di immagini (Shutterstock, Getty). Compensano gli autori in base alla quantità di contenuti forniti, ma ignorano la qualità o l'utilità delle opere. Startup e ricercatori stanno integrando questo modello per garantire una distribuzione più equa basata sulla similarità. Offriamo al lettore un simulatore online che gli permetterà di verificare personalmente, su scala reale, la fattibilità tecnica ed economica di questo approccio.

In prospettiva, questo modello protegge i diritti degli autori e si oppone all'“Uberizzazione dell'intelligenza artificiale”. La “sacemizzazione” valorizza la “proprietà liquida”, la ricchezza intellettuale della creatività, che mantiene e sviluppa il capitale culturale e umano di fronte alla sua estrazione da parte dei monopoli.

Raccomandiamo pertanto a Sam Altman e agli operatori dell'intelligenza artificiale generativa un modello economico in collaborazione con le società degli autori e a tutte le parti interessate un metodo per avviare trattative in questo nuovo quadro di riflessione.

**Sfide**

# 1 – Chi vuole che l'intelligenza artificiale generativa paghi gli autori?

"Penso che abbiamo bisogno di un nuovo accordo, un nuovo standard, un nuovo protocollo, come vuoi chiamarlo, per stabilire come verranno ricompensati i creatori", ha affermato Sam Altman, fondatore di Open AI. "Abbiamo bisogno di nuovi modelli economici che consentano ai creatori di generare reddito. »

L'idea di pagare gli autori tramite gen-AI è stata sostenuta da altre personalità per diversi mesi.

Operatori di intelligenza artificiale che vogliono pagare gli autori

NOME	AZIENDA	CITAZIONE
sam altman	AI aperta	"Penso che abbiamo bisogno di un nuovo accordo, un nuovo standard, un nuovo protocollo, come vuoi chiamarlo, per determinare come verranno ricompensati i creatori. Abbiamo bisogno di nuovi modelli economici che consentano ai creatori di generare reddito. »
Arthur Mensch	Mistral AI	"Abbiamo interesse ad avere accesso a contenuti di qualità. Abbiamo quindi avviato discussioni con i fornitori di contenuti, nell'editoria e nella stampa. Si possono trovare sinergie e meccanismi di condivisione del valore. »
Davide Holz	A metà viaggio	"Non esiste un modo concreto per prendere cento milioni di immagini e capire da dove provengono. Sarebbe bello se le immagini avessero dei metadati incorporati sul proprietario del copyright o qualcosa del genere. Ma questo non è possibile, non esiste alcun registro. Non esiste un modo per trovare un'immagine su Internet e poi risalire automaticamente al suo proprietario, per poi poter fare qualcosa per autenticarla. »
Chris Dixon	a16z	"Il movimento [Web 3] ha il potenziale per riportare in auge lo spirito dell'Internet delle origini, garantire i diritti di proprietà dei creatori, rivendicare la proprietà e il controllo degli utenti e spezzare la morsa delle grandi aziende che incentrano le nostre vite. "
Autore: Neal Stephenson	Lamina1 (inventore del metaverso)	"È possibile monitorare quanta influenza ha avuto una determinata immagine o un dato input sul risultato finale. E poi, se il risultato finale vale qualcosa, allora forse c'è un modo per dare credito e una compensazione proporzionale per quegli input [...] Diciamo che stavo scrivendo un libro che conteneva una spada magica appena descritta nel libro. [...] Qualcuno potrebbe quindi creare un bene che potrebbe essere venduto sull'Unreal Asset Store [...] Ci sono dei sound designer che devono fare qualcosa di simile con i suoni che produce quando viene utilizzato. E ci sono programmatori che devono integrare la spada nel gioco, utilizzando progetti o C++, in modo che sia effettivamente in grado di fare qualcosa e contribuire all'esperienza. Quindi a ogni passaggio viene aggiunto più valore. E alla fine di questo processo, avrai qualcosa che potrebbe effettivamente generare entrate. E quando finalmente arrivano queste entrate, quello che vorresti che facessero è propagarsi all'indietro. E vorresti che le diverse persone che hanno contribuito alla catena del valore fossero in qualche modo compensate"
Jaron Lanier	Microsoft	"Le persone saranno pagate per i loro dati e pagheranno per servizi che richiedono dati da altre persone. L'attenzione degli individui sarà guidata dai loro interessi autodefiniti piuttosto che da piattaforme manipolative dipendenti da inserzionisti o altre terze parti. Le piattaforme riceveranno dati di qualità superiore con cui addestrare i loro sistemi di apprendimento automatico. » Lanier riconosce che persino i ricercatori sulla dignità dei dati non riescono a mettersi d'accordo su come districare tutto ciò che i modelli di intelligenza artificiale hanno assorbito o su quanti dettagli siano necessari per tenerne conto. Eppure Lanier pensa che si potrebbe fare, anche se gradualmente. "Dobbiamo calcolare e presentare la provenienza delle fonti umane più importanti per un dato risultato di intelligenza artificiale. Al momento non lo facciamo. Possiamo farlo in modo efficace ed efficiente, semplicemente non ci riusciamo. Deve essere una decisione sociale quella di fare questo passo."

NOME	AZIENDA	CITAZIONE
Yat- Siu (Si)	Marchi Animoca	"I dati sono la risorsa più preziosa. Innanzitutto, sono alla base di tutta l'intelligenza artificiale di cui sentiamo tanto parlare. Nessun dato, nessun ChatGPT . I dati sono il nuovo lavoro. E non veniamo compensati equamente per questo. [...] Esiste una chiara correlazione tra i diritti di proprietà e la ricchezza delle nazioni. » "Immaginate di andare su Facebook e alla fine della giornata di realizzare di aver appena fatto guadagnare mille dollari all'azienda. Il tuo rapporto con Facebook cambierebbe completamente. Probabilmente pretenderesti la tua giusta quota. »
Yann Le Cun	Meta	"un modello di propagazione dei diritti [di copyright] in una direzione e dei ricavi nell'altra"
Viaggio Adler	ex CEO di Scribd	"La quarta legge è un insieme di principi guida su come le aziende di intelligenza artificiale possono utilizzare e formare contenuti generati dagli esseri umani. La quarta legge, ispirata alle Tre leggi dei robot dell'autore di fantascienza Isaac Asimov, afferma che gli esseri umani dovrebbero avere il diritto di acconsentire e controllare il modo in cui l'intelligenza artificiale utilizza il loro lavoro, di essere compensati (se richiesto) e accreditati per il loro lavoro (se un libro è menzionato nel risultato, dovrebbe esserci un link per acquistarlo). »

Con la leadership di Sam Altman, l'idea di risarcire gli autori assume maggiore consistenza, anche se egli ritiene che tale risarcimento sarebbe dovuto solo nel caso in cui lo stile dell'autore venisse plagiato, il che è molto limitante.

Mette però in discussione il modello economico della sua attuazione . Lo scopo di questa nota è quello di proporre questo nuovo modello economico nel quadro della legge sull'intelligenza artificiale .

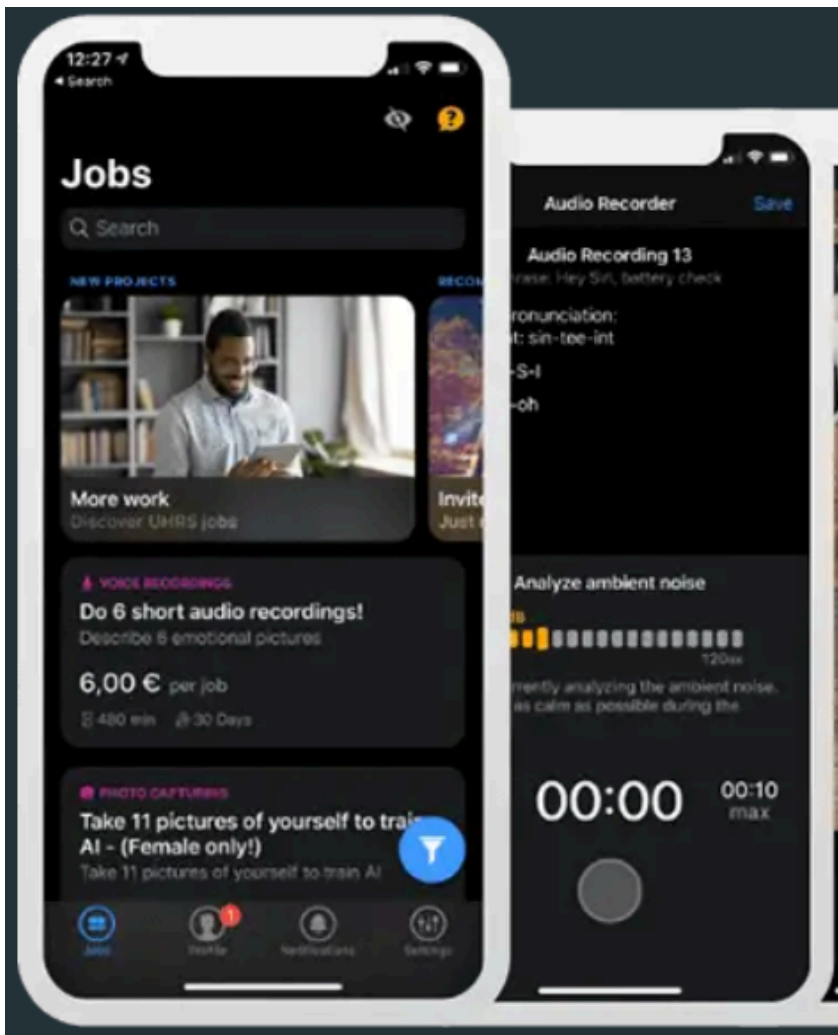
## 2 – Le tre facce della remunerazione degli autori

Se affrontato a livello puramente legale, il conflitto sullo scraping , che contrappone autori e operatori dell'intelligenza artificiale generativa, durerà almeno dieci anni.

Nel frattempo, il settore si segmenterà in creatori di dati, piattaforme, aggregatori, negozianti, formatori, raffinatori, sfruttatori e integratori di intelligenza artificiale generativa, il che diluirà l'atto della creazione. Il patrimonio culturale sarà stato mescolato, diffuso e riciclato in contenuti le cui tracce originali saranno state cancellate.

Per paura di una degenerazione dei contenuti didattici, si sta già sviluppando un nuovo settore di micro-attività per la produzione di massa di contenuti.





©Clickworker 2024

della gig economy per l'intelligenza artificiale si basano sulla logica dei comandi. Clickworker paga direttamente 6 € agli utenti per descrivere ad alta voce 6 immagini e pubblicare la registrazione. Oppure scattarti una foto mentre fai esercizi fisici. Gigwalk e Premise pagano 4 dollari per scattare una foto di un ingorgo a un incrocio. Il risultato viene utilizzato per addestrare le IA.

Questo modello di tipo “Uber dei dati” sta crescendo rapidamente grazie all’enorme necessità di dati da parte degli operatori di intelligenza artificiale generativa.

Questo è l'opposto delle società di copyright come la SACEM, che si basano sulla fornitura e sulla proprietà intellettuale degli autori, che monitorano questa proprietà, riscuotono le royalty e remunerano i titolari dei

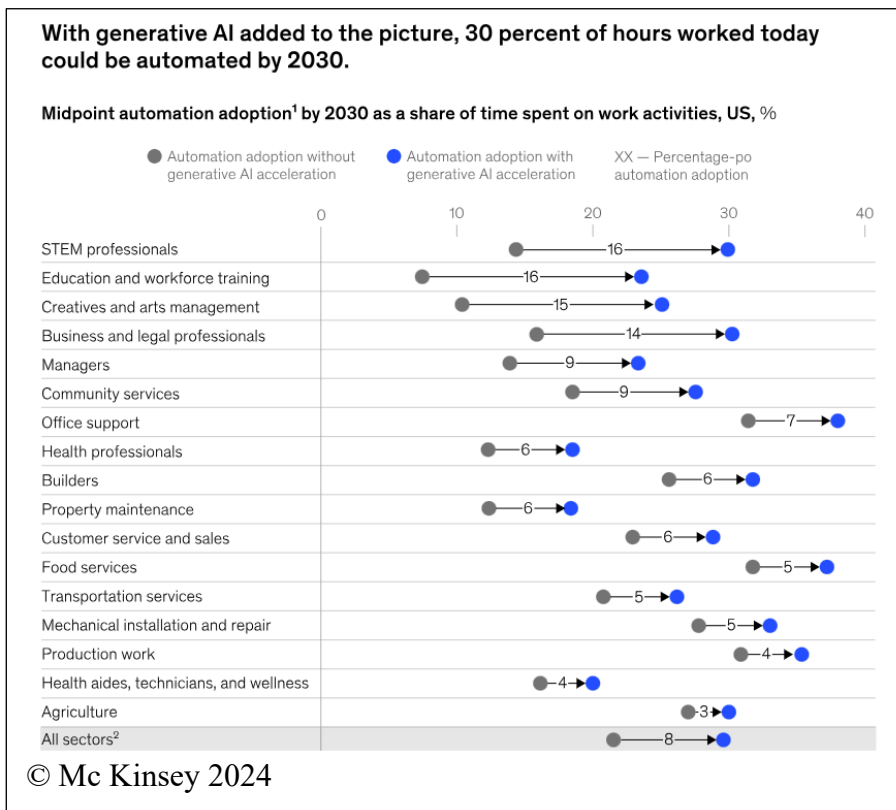
diritti secondo scale di distribuzione predefinite.

Il primo modello privilegia la domanda, l'esecuzione, la mercificazione immediata, il neo-taylorismo cognitivo, la standardizzazione del lavoro. Il secondo promuove l'offerta, l'iniziativa, la proprietà intellettuale, la creatività, la diversità culturale. Il primo sfrutta il capitale culturale, il secondo lo coltiva.

Tra gli altri modelli economici (vedi sotto), scraping, data Uber e le società d'autore sono i tre principali volti della remunerazione – o meno – degli autori, il cui squilibrio aggrava le tensioni attuali.



### 3 – Oltre il copyright, il capitale umano



Di fronte allo sviluppo dell'intelligenza artificiale generativa, autori e creatività sono in prima linea. Seguiranno l'artigianato, le professioni artistiche, la competenza, la ricerca e tutte le professioni dotate di know-how, in tutti i settori economici.

Man mano che il modello “Uber dei dati” si diffonderà, darà origine a una forma di intelligenza artificiale “Uberizzante”, nello stesso modo in cui l'azienda Uber un tempo diede inizio al fenomeno dell'Uberizzazione.

Ad esempio, la personalizzazione dell'intelligenza artificiale generativa aziendale con tutti i

suoi archivi dati consente di ottenere una bozza per ogni nuovo messaggio, report, progetto, presentazione e prezzo. Le correzioni professionali addestrano l'IA fino a quando il tasso di errore diventa accettabile. Gli incrementi di produttività sono significativi ma, in cambio, le posizioni dell'azienda sui collegamenti strategici di creazione di valore si stanno riducendo a vantaggio dell'operatore di intelligenza artificiale.

Poiché tutte queste competenze sono l'intersezione di conoscenze, pratiche e materiali di organizzazione sociale, è il tessuto delle aziende ad essere in gioco. È come se, al tempo della rivoluzione industriale, un cartello di proprietari di motori a vapore controllasse l'organizzazione taylorista del lavoro. Ecco cosa si aspetta il mercato azionario dalle Big Tech. Concedendo capitalizzazioni nell'ordine dei trilioni, non valorizza più le aziende, ma i monopoli.

L'idea che le macchine abbiano il “diritto di imparare” gratuitamente si diffonderà. Tra dieci anni, anche se gli autori vincessero la battaglia legale contro lo scraping, il loro compenso si baserebbe sui dati comparabili forniti da Uber, il che non è molto. Col senno di poi, ci renderemo conto che il diritto d'autore esistente non era altro che una linea Maginot facile da aggirare.

L'economia dispone di capitale culturale e umano, da tempo individuato come una risorsa considerevole, al punto da essere posto al centro della strategia economica europea. Questa strategia, troppo precoce, fu infine dimenticata a causa della mancanza di un modello economico. Sottovalutato e mal difeso, questo capitale è ora minacciato dall'estrazione da parte dell'intelligenza artificiale “Uber”. Per ogni azienda, paese o continente, rifiutare la tecnologia significa condannarsi a un declino duraturo, come quello della Cina durato due secoli. Accettarlo senza compensazione significa lasciarsi colonizzare senza combattere.

Il premio Nobel Jean Tirole ci insegna che 1/ la ricchezza delle nazioni dipende dalla loro capacità di catturare valore a livello di innovazione. 2/ L'innovazione nasce dall'iniziativa individuale, perché la caratteristica del lavoro creativo è che non sappiamo cosa troveremo. 3/ La proprietà intellettuale è un male necessario per stimolare questa creatività. 4/ impedire a un'azienda di commercializzare i propri prodotti finché non ha pagato le royalties è un'arma molto efficace contro Gafam . 5/ I pool di proprietà intellettuale con autorizzazione di licenza individuale richiederebbero loro di acquisire le licenze, ma a un prezzo competitivo.

Le nazioni che vogliono migliorare il loro capitale umano e culturale devono trovare una soluzione moderna di sacemizzazione , il cui modello economico possa competere efficacemente con l'intelligenza artificiale " uberizzante " .

# Soluzione tecnica

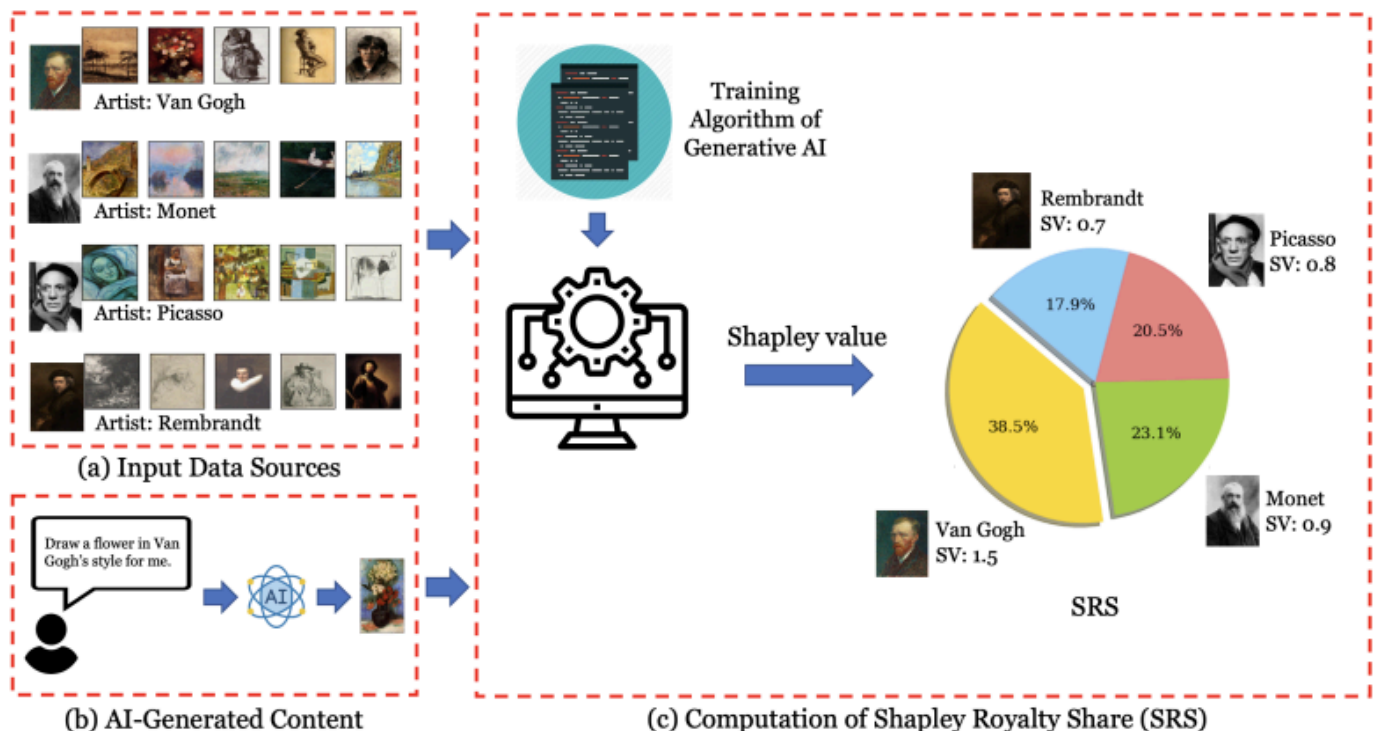
## 4 - Soluzione ideale

Nella teoria dei giochi, Lloyd Shapley introdusse nel 1953 l'idea che fosse possibile, in un gioco cooperativo, calcolare un'equa distribuzione dei guadagni tra i giocatori. Ad esempio, consideriamo un ecosistema marino con tutte le sue specie viventi (alghe, anemoni, conchiglie, stelle marine) e misuriamo l'utilità di ciascuna specie in base al suo contributo alla vita nell'ecosistema. Quando si eliminano le stelle marine, scompare il 50% delle altre specie, rispetto al 10% che si verifica quando si eliminano gli anemoni. Il valore Shapley delle stelle marine è cinque volte quello degli anemoni.

Esiste un consenso tra i ricercatori sulla soluzione ideale per la remunerazione degli autori. Ciò comporta il calcolo del valore Shapley per ciascun autore e la distribuzione della remunerazione in proporzione a tale valore.

Nel caso dell'intelligenza artificiale generativa, l'equa distribuzione della remunerazione per i lavori di formazione dovrebbe idealmente essere effettuata in base all'utilità dei lavori di formazione per i contenuti generati dall'intelligenza artificiale, da parte dell'autore. Nella figura sottostante, il valore Shapley di Van Gogh è pari a 1,5, ovvero al 38,5% della retribuzione totale.

Valore Shapley



Nella pratica, il costo del calcolo del valore Shapley è proibitivo, il che solleva la questione della fattibilità tecnica di una soluzione di remunerazione equa.

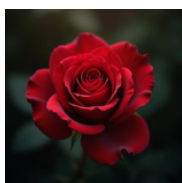
## 5 – Dimostratore

### Oggetto

Il lettore troverà un simulatore online su <https://huggingface.co/spaces/TheFrenchDemos/genai-rew-a-rd>








Riteniamo che ora sia tecnicamente ed economicamente possibile calcolare un punteggio di attribuzione per i contenuti formativi, al fine di distribuire equamente la remunerazione tra i titolari dei diritti di tali contenuti. Questo simulatore permetterà a tutti di verificare autonomamente le proprie ipotesi economiche.

Ad esempio, è stata generata un'immagine di una rosa con il prompt " *Una rosa rossa vibrante in piena fioritura, fotografia macro* ":



Il simulatore identifica in pochi secondi, tra i dieci milioni di contenuti del database Open Images, le otto immagini originali che hanno contribuito maggiormente alla formazione del contenuto generato. Calcola un tasso di attribuzione, un compenso e fornisce informazioni identificative sul contenuto originale e sul suo autore.

### Attribuzioni delle immagini di rose

CONTENUTO	ATTRIBUZIONE	RICOMPENSA CONTENUTO (c€)	URL DEL CONTENUTO	NOME DELL'AUTORE	URL DELL'AUTORE
	21%	0,63	Italiano: <a href="https://www.flickr.com/photos/50093642@N03/4629227563">https://www.flickr.com/photos/50093642@N03/4629227563</a>	Yoko Nekonomania	Italiano: <a href="https://www.flickr.com/photos/nekonomania/">https://www.flickr.com/photos/nekonomania/</a>
	18%	0,54	Italiano: <a href="https://www.flickr.com/photos/22305657@N02/11733854773tt">https://www.flickr.com/photos/22305657@N02/11733854773tt</a>	Bruno Pinho	Italiano: <a href="https://www.flickr.com/photos/brunopinho/">https://www.flickr.com/photos/brunopinho/</a>
	15%	0,45	Italiano: <a href="https://www.flickr.com/photos/12692992@N00/170278141">https://www.flickr.com/photos/12692992@N00/170278141</a>	Decano + Barb	Italiano: <a href="https://www.flickr.com/photos/barbanddean/">https://www.flickr.com/photos/barbanddean/</a>
	12%	0,36	Italiano: <a href="https://www.flickr.com/photos/33140999@N07/3509399376">https://www.flickr.com/photos/33140999@N07/3509399376</a>	Michele Arrighi	<a href="https://www.flickr.com/photos/arrighi/">https://www.flickr.com/photos/arrighi/</a>
	10%	0,30	Italiano: <a href="https://www.flickr.com/photos/55658968@N00/8730223835">https://www.flickr.com/photos/55658968@N00/8730223835</a>	Toshihiro Gamo	Italiano: <a href="https://www.flickr.com/photos/dakiny/">https://www.flickr.com/photos/dakiny/</a>
	9%	0,27	Italiano: <a href="https://www.flickr.com/photos/34226407@N04/8734050294">https://www.flickr.com/photos/34226407@N04/8734050294</a>	Principe Edd	Italiano: <a href="https://www.flickr.com/photos/edpics/">https://www.flickr.com/photos/edpics/</a>
	9%	0,27	Italiano: <a href="https://www.flickr.com/photos/41353201@N07/4267426012">https://www.flickr.com/photos/41353201@N07/4267426012</a>	Musicista Loxton	Italiano: <a href="https://www.flickr.com/photos/dj-dwayne/">https://www.flickr.com/photos/dj-dwayne/</a>

	8%	0,24	Italiano: <a href="https://www.flickr.com/photos/64574638@N04/6006728032">https://www.flickr.com/photos/64574638@N04/6006728032</a>	serbianstruhpro	Italiano: <a href="https://www.flickr.com/photos/serbestruhpro/">https://www.flickr.com/photos/serbestruhpro/</a>
---	----	------	--	-----------------	--

Nella parte inferiore della pagina web, tre pulsanti consentono all'utente di scaricare:

- Questo rapporto
- La relazione dei titolari dei diritti che l'operatore potrebbe generare e comunicare ai titolari dei diritti in applicazione degli accordi transattivi e della retrocessione dell'equa remunerazione
- La relazione statutaria prevista dalla legge AI (art. 53, 1, d), secondo il modello proposto dal Consiglio superiore per l'arte letteraria e artistica Proprietà (CSPLA)

## Come funziona

Il simulatore ha un pulsante " avanzato " parametri » che consente all'utente di regolare i parametri desiderati. Nella tabella seguente sono riportate quelle impostate di default e sono indicate le formule di calcolo.

Parametri e valori di una simulazione

VARIABILE	COLLOCAMENTO	COMPARABILI	CALCOLO	FORMULA
Abbonamento utente	12 €/mese (1)	Adobe Firefly a partire da \$ 4,99/mese Midjourney da \$ 10/mese Dall-e incluso con ChatGPT Plus \$20 Getty Edify 45 € per 25 generazioni	S	
Contenuto generato	60 /mese	Adobe Firefly Basic Plan 100/mese Piano base Midjourney 100/mese	G	
Valore nominale per contenuto generato			0,2 €	$V = S / G$
Quota di titolari di diritti e autori				$R\% = Rr\% + Ra\%$
- Titolari dei diritti	10%	Musica (ASCAP, BMI, SACEM) 2 – 8% Immagine stock 55 – 85%		$\% Rr$
- Autori	5%	Libri stampati 5 – 15% Musica (esecutore e compositore) 15 – 30% Fotografia stock 15 – 45%		$Ra\%$

Remunerazione dei titolari dei diritti e degli autori per i contenuti generati	3 centesimi €	$R = V \times R\%$
Attribuzione	21%	$A\% = \text{Algoritmo di similarità}$ $N = \text{numero di contenuti assegnati}$
Ricompensa per i contenuti di formazione	0,63 centesimi €	$T = R \times A\%$
Numero di autori	1.800	B
Utenti	500.000	Si stima che il numero di abbonati paganti di Midjourney sia compreso tra 2 e 5 milioni . Io
Ricompensa mensile per il titolare dei diritti e l'autore	829 €	$S \times U \times R\% / B \times A\% \times N$
- Titolari dei diritti	276 €	$S \times U \times Rr\% / B \times A\% \times N$
- Autore	553 €	$S \times U \times Ra\% / B \times A\% \times N$

## Concetto di somiglianza

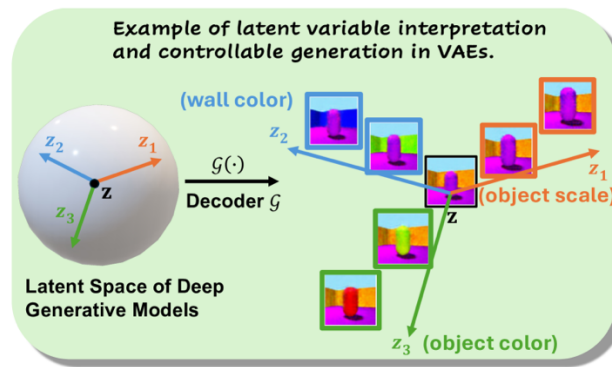
Questo dimostratore utilizza un algoritmo di similarità tra il contenuto di formazione e il contenuto generato, implementando [DINOV2](#) e [FAISS](#). Intuitivamente, due immagini simili si assomigliano.

Tuttavia , la "somiglianza" qui menzionata è una nozione scientifica slegata da qualsiasi nozione giuridica di "contraffazione", "somiglianza" o contraffazione di un'opera, definita nel diritto francese come "qualsiasi riproduzione, rappresentazione o distribuzione, con qualsiasi mezzo". "qualunque sia il significato di un'opera intellettuale in violazione dei diritti dell'autore, come definiti e regolati dalla legge" (CPI, art. L. 335-3). Non vi è necessariamente violazione o violazione mediante la riproduzione di un elemento di un'opera, un "contributo" in questo caso. Allo stesso modo, quando un autore crea, un elemento può entrare nella sua composizione, senza che vi sia alcuna contraffazione o violazione da parte di terzi. La remunerazione derivante dalla "somiglianza" non dovrebbe quindi paralizzare l'azione dell'autore in caso di violazione.

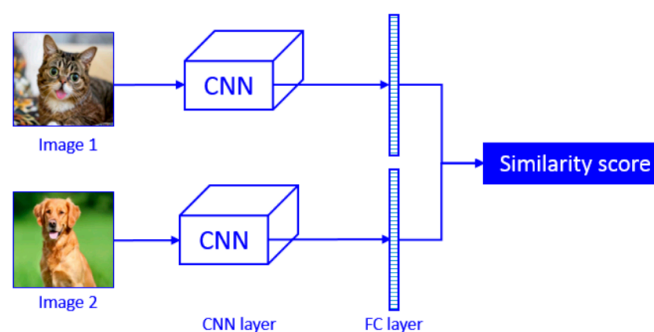
## Somiglianza vettoriale

Tecnicamente, ogni contenuto generato o originale viene analizzato per estrarre informazioni misurabili: gatto, cane, colore e dimensione del primo piano, tipo, stile, dimensione, oggetti riconosciuti, parole chiave, ... tutti questi centinaia di indicatori di utilità , circa 500 nel nostro dimostratore , forma un vettore in uno "spazio latente".





Il grado di attribuzione di un'opera (immagine 1) a un contenuto generato (immagine 2) viene calcolato tramite un punteggio di similarità tra le due immagini, elaborato dalle due reti neurali siamesi. Quanto più vicini sono i vettori risultanti, tanto più alto è il punteggio di similarità tra le immagini.



## Numero di autori

Il database Open Images è stato creato a partire da quello di Flickr, applicando filtri quali la non rilevazione da parte di Google Image Search, un numero minimo di pixel o la deduplicazione. Open Images seleziona anche solo le licenze CC-BY, ovvero un'immagine su 30. Poiché Flickr offre di default la licenza "tutti i diritti riservati", la modifica della licenza viene effettuata solo da autori informati. Ciò suggerisce un'elevata concentrazione in Open Images dei collaboratori più attivi, che hanno pubblicato più di 5.623 foto, ovvero circa 1.800 autori. Questa ipotesi intuitiva potrebbe essere verificata con un conteggio esatto degli autori del database Open Images, cosa che non siamo stati in grado di ottenere a causa del limite nella velocità di query.

## Identificazione degli autori

Il database Open Images identifica gli autori delle immagini e fornisce la loro pagina del profilo Flickr, che tuttavia non include un indirizzo e-mail o postale completo.

Oltre al nostro dimostratore e alle Open Images, i set di dati pubblici forniscono elementi identificativi, di precisione variabile, ma che possono essere incrociati per identificare ciascun titolare dei diritti.

I metodi di riferimento incrociato sono noti soprattutto per il loro uso fraudolento contro la privacy. Già nel 1997, la comunicazione obbligatoria di statistiche aggregate da parte di una compagnia assicurativa del Massachusetts, insieme ai dati elettorali, aveva reso possibile l'identificazione senza possibilità di errore del governatore William Weld e la conoscenza della sua cartella clinica.

Il loro utilizzo legale consente, nella maggior parte dei casi, la completa identificazione degli autori: tramite il confronto dei grafici sociali, tramite la correlazione statistica grazie a una rara combinazione di attributi (nicchia scientifica, ore di attività, indirizzo IP), tramite la ricostruzione dei singoli attributi dalle medie di gruppo e dagli attributi noti di altri individui, ecc. La tabella seguente fornisce indizi per identificare gli autori dei contenuti formativi GPT 4.



AI-generated by Adobe Firefly

Possibile identificazione degli autori dei set di dati utilizzati da GPT-4

	SET DI DATI	ELEMENTI DI IDENTIFICAZIONE	INCROCIO DIRETTO	IDENTIFICATORE
<b>Identificazione diretta</b>	Immagini aperte	-	-	Nome, cognome, scheda profilo
	Progetto Gutenberg	-	-	Nome, cognome, data di nascita e di morte
	arXiv	-	-	Nome, Iniziale_Nome
	Pubblicazione	-	-	Id_autore
<b>Identificazione indiretta</b>	ComuneCrawl	Dominio	Chi è	Registrant_contact (nome, organizzazione, via, città, stato, telefono)
	LibroCorpus	Testo normale	Parole di distruzione	Nome, cognome
	Testo Web	Plain_text + search = "scritto da"	API di Amazon	Nome, cognome
	Sovraccarico di pila	Codice HTML	Italiano:	ID_utente , profilo_di_rete
<b>In attesa di richiesta</b>	Wikipedia	Nome utente	-	-
	Reddit	Nome utente	-	-
<b>Altro</b>	Voce comune	(statistiche)	-	-
	Google Libri Ngram	( non applicabile )	( n / a )	( n / a )
	Guida in linea	( non accessibile )		

## 6 – Soluzioni esistenti

### Sintesi

La tabella seguente identifica alcuni metodi commercializzati o testati, relativamente al nostro dimostratore:

Metodi per promuovere i contenuti dell'autore

VALUTAZIONE	METODO DI VALUTAZIONE	DIMOSTRATORE	SOCIETÀ COMMERCIALI	RICERCATORI
Contenuto generato	Costo/quantità dell'abbonamento	Controversia		
Insieme di dati	Somiglianza dei set di contenuti Il signor Shapley in esecuzione Shapley Frequenza di accesso al sito web, tipo di dati		Pedaggiobit	Adobe AI aperta AI aperta
Lavoro di formazione	Somiglianza del contenuto Modellazione dell'apprendimento Personalizzazione del modello disimparare Tatuaggio Shapley federato	Controversia	Bria (immagini), Prorata Gist.ai (testi)	AI aperta Adobe Adobe Adobe Huawei
Autore	Pagare per allenarsi Influenza dell'intera opera Identificazione dei titolari dei diritti		Shutterstock , Getty , Audiosparx , Canva Sureel.ai (musica)	
		Controversia		

## Soluzioni commercializzate

Aziende che pagano gli autori per i contenuti formativi

AZIENDA	CAPITALE DI RISCHIO	FATTURATO	METODO	MEDIA	FORNITORE DI CONTENUTI
Adobe (Firefly)	-	21 miliardi di dollari	Pagare per allenarsi	Immagine	
Tela	-	2 miliardi di dollari	Pagare per allenarsi	Immagine	
Getty	-	917 milioni di dollari	Pagare per allenarsi	Immagine	
Immagine	-	875 milioni di dollari	Pagare per allenarsi	Immagine	
Prorata (Gist.ai)	25 milioni di dollari	-	Somiglianza dei contenuti	Testo	Sky News, Tempo, Financial Times, The Atlantic, Axel Springer, The Guardian, dmg media, Universal Music, Fortune
Bria	24 milioni di dollari	-	Somiglianza dei contenuti	Immagine	Shutterstock, Getty Images, Alamy , Superstock
Pedagogibit	24 milioni di dollari	-	Frequenza di accesso al sito web	Testo	Penske Media Corporation, Tempo, Mumsnet, Marchi di media affidabili, Candr Media Group, AdWeek
Sureel.ai	-	-	Influenza dell'intera opera	Immagine, musica	
Audiosparx	-	-	Pagare per allenarsi	Musica	

## Metodo di pagamento per l'addestramento

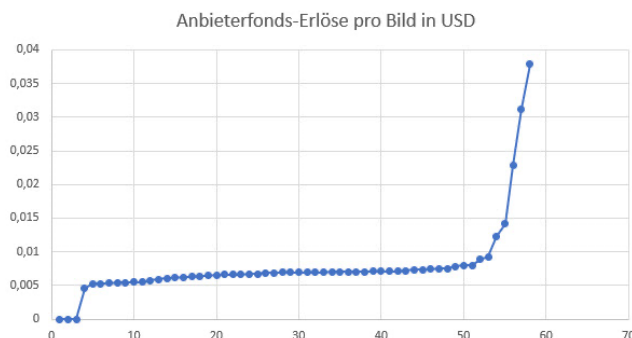


AI-generated by DALL-e

Le banche di contenuti private di Adobe, Canva , Getty e Shutterstock , distribuiscono il prezzo dei set di dati che vendono utilizzando il metodo " pay to train". Gli autori vengono compensati in base al volume dei loro contributi a questi set di dati .

Il valore “ pay -to-train” può essere calcolato da vari indicatori di utilità. Shutterstock, ad esempio, paga 0,0050 dollari all'anno per foto, fino a 0,0038 dollari all'anno per foto, a seconda della quantità di metadati e di altre metriche.

Valore del pay -to-train : remunerazione degli autori da parte di Shutterstock . Campione N = 58 (in \$/immagine)



Il vantaggio di questo modello è la semplicità di calcolo. È operativo da due anni e sembra essere ben accolto dai titolari dei diritti.

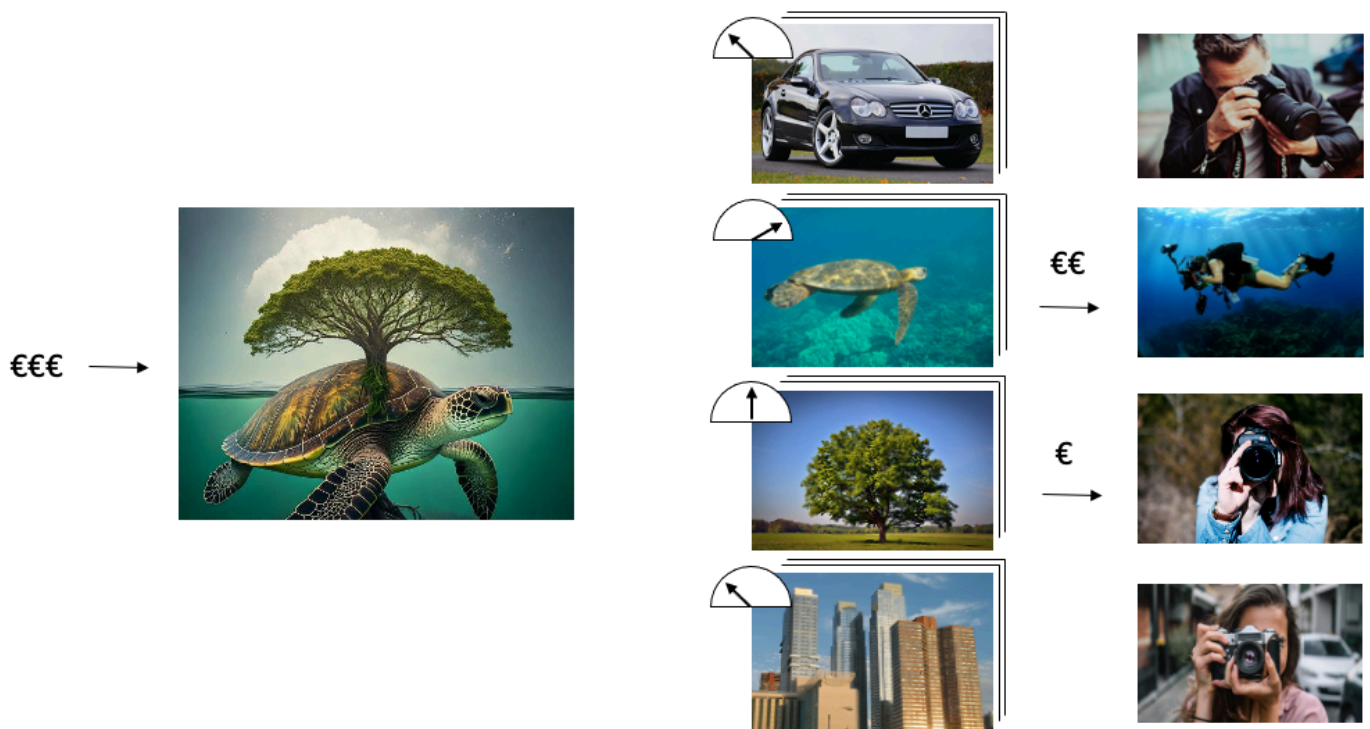
D'altro canto, il livello di equità tra gli autori presenta due distorsioni:

- La qualità del contenuto non viene presa in considerazione: un autore che ha contribuito con 200 foto riceve, a parità di altre condizioni, il doppio della remunerazione di un collaboratore che ne ha contribuite con 100.
- l'utilità del contenuto generato non viene presa in considerazione: Van Gogh non riceve più di Picasso se gli utenti inseriscono più spesso "alla maniera di Van Gogh" nei loro prompt.

## Metodo ibrido Similarità + Pay -to-train

La startup israelo -americana Bria, che collabora con grandi banche di immagini private ( Shutterstock , Getty), è integrata da Big Tech (Microsoft, Amazon, Nvidia ). Come primo approccio, Bria utilizza, come il nostro dimostratore, un metodo di similarità, per remunerare gli autori proporzionalmente alla somiglianza così misurata.

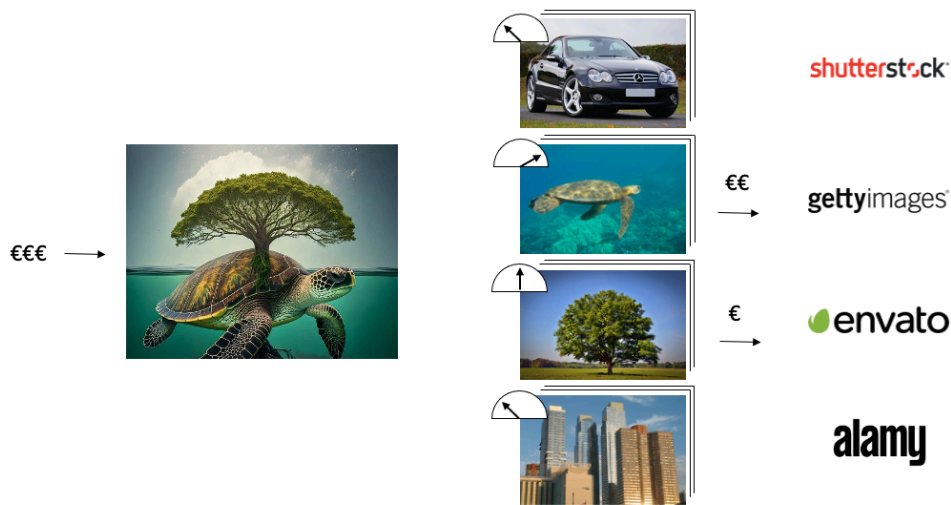
Misurare le somiglianze tra i contenuti generati e le foto originali per remunerare gli autori



Più precisamente, pare che la somma dei punteggi di similarità venga calcolata dalle banche di immagini ( Shutterstock , Getty, Alamy ...), il che determina una remunerazione iniziale delle banche proporzionale a queste somme.

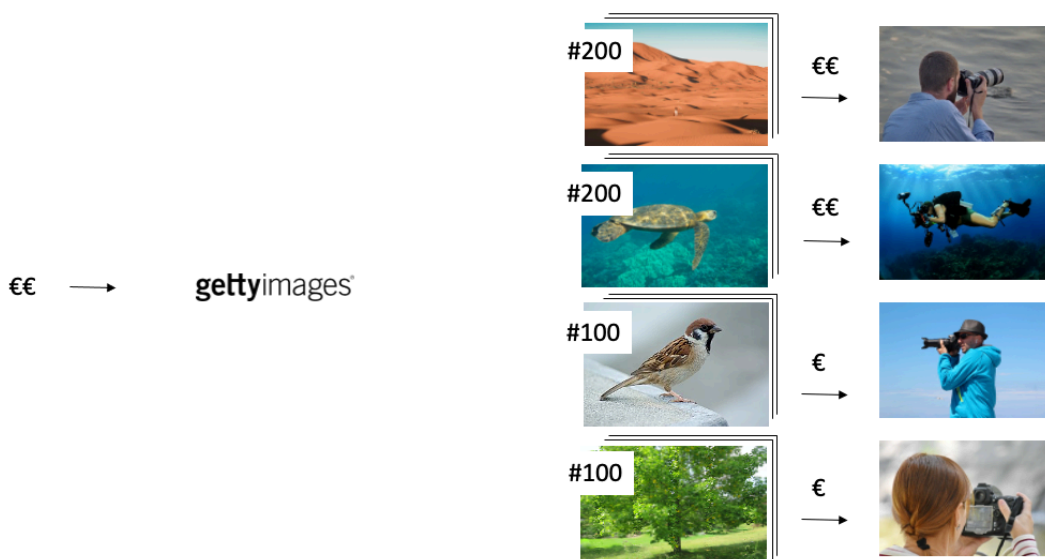


Misurare le somiglianze tra i contenuti generati e le foto originali per remunerare le banche di immagini



Queste banche distribuiscono poi questa prima remunerazione tra gli autori, in base al numero individuale delle loro opere che contribuiscono ai set di dati venduti, secondo il metodo “pay to train” .

Numero di foto da pagare agli autori



Vengono apportate ulteriori migliorie e le immagini con metadati di qualità ricevono un bonus.

Il metodo di Bria si svolge quindi in due fasi:

- Innanzitutto, grazie ad un algoritmo di similarità tra le immagini di training e le immagini generate
- Quindi, grazie ad un algoritmo di contributo quantitativo degli autori ai dataset venduti (“pay -to-train”)

Questa soluzione ibrida è:

- economicamente compatibile con il modello di business dell'intelligenza artificiale generativa

- accettato per due anni da banche di immagini e autori
- attualmente esteso ad altri tipi di contenuti: 3D, musica, voce, immagini animate, video

È possibile che in particolare la startup ProRata abbia adottato questo metodo ibrido e lo offra ai media che hanno già accordi con gli operatori di intelligenza artificiale. Si tratta di “prendere l’output dell’intelligenza artificiale generativa, che si tratti di testo, un’immagine, musica o un film, e scomporlo nei suoi componenti, per capire da dove provengono, quindi assegnare una percentuale a ciascun detentore del copyright e pagarli di conseguenza .” I brevetti pubblici della società madre non forniscono ulteriori dettagli.

Questo metodo ibrido presenta ancora alcuni difetti:

- Non è dimostrato che questa soluzione si adatti a tutti i tipi di contenuti, ad esempio i giochi
- non vi è alcuna garanzia che la somiglianza non sia casuale. In altre parole, un autore potrebbe essere pagato anche se le sue opere non fossero utili per il contenuto generato, e viceversa.
- il grado di contributo di un'immagine originale viene misurato approssimativamente. Nell'esempio sopra, il fotografo del deserto viene pagato quanto il fotografo delle tartarughe marine perché il numero dei loro contributi al set di dati Getty Images è lo stesso, mentre le foto del deserto non hanno alcuna utilità per il contenuto generato.

Va notato che:

- anche se la distribuzione del compenso per un particolare contenuto generato è approssimativa, la sua equità probabilmente migliora con la quantità di contenuto generato.
- non è necessaria l'identificazione individuale degli autori da parte di Bria, ma solo l'identificazione dei set di dati di formazione
- l'algoritmo di distribuzione in sé non è pubblico. È possibile che in futuro gli autori richiederanno maggiore trasparenza, ma è anche probabile che, al contrario, sarà necessaria una certa dose di segretezza per evitare pregiudizi opportunistici.
- Non conosciamo inoltre il tasso medio di remunerazione degli autori in rapporto al fatturato dei contenuti generati che, a nostro avviso, potrebbero essere pubblicati.

## Metodo di similarità del set di dati

La startup Sureel fornisce ai musicisti uno strumento che misura il livello di influenza del loro lavoro, espresso in percentuale, su una traccia generata dall'intelligenza artificiale.

L'algoritmo, descritto nel suo brevetto del '949, prevede il compenso di un autore in base all'influenza di tutte le sue opere, misurata ad esempio dalla differenza tra i pesi del modello base e quelli della sua regolazione fine da parte di questo insieme. Questa variazione mira a diluire il pregiudizio di somiglianza attraverso il confronto di un numero maggiore di opere.



## Metodi attualmente in fase di sperimentazione

Oltre a queste poche varianti, i ricercatori stanno gareggiando in ingegnosità per testare nuovi metodi:

Ricerca di modelli di compensazione equa

ATTIVITÀ COMMERCIALE	METODO	TEAM DI RICERCA	DATA	MOTIVAZIONE
AI APERTA Università di Princeton Università della Columbia Università di Harvard Università della Pennsylvania Università della California Berkeley	Il signor Shapley	<b>Una soluzione economica alle sfide del copyright dell'intelligenza artificiale generativa</b> Jiachen T. Wang, Zhun Deng, Hiroaki Chiba-Okabe, Boaz Barak, Weijie J. Su	14 aprile 2024 9 settembre 2024 (v4)	I valori Shapley forniscono un approccio basato su principi per la distribuzione dei guadagni in base all'utilità.  L'utilità aiuta a misurare in che misura le fonti di dati sono responsabili della generazione di contenuti. Il valore è basso se è improbabile che il modello controfattuale generi lo stesso contenuto del modello distribuito e viceversa.  L'utilità può essere interpretata come la remunerazione totale che tutti i membri meritano collettivamente per aver fornito i propri dati per addestrare il modello di intelligenza artificiale generativa.
Tecnologia della Virginia Istituto di tecnologia del Massachusetts	Dati in esecuzione Shapley	<b>Data Shapley in un'unica sessione di allenamento</b>  Jiachen T. Wang, Prateek Mittal, Dawn Song e Ruoxi Jia	29 giugno 2024	Boaz Barak, professore di sicurezza ad Harvard ("superallineamento") e collaboratore di OpenAI, non perde di vista il metodo ideale di Shapley, anche se ciò significa scendere a compromessi in termini di accuratezza.  Il suo team di ricercatori dimostra nella pratica come calcolare il valore Shapley di set di dati come Wikipedia o Github per un modello di intelligenza artificiale di uso generale addestrato a partire da un set di dati di tipo GPT.  Secondo questo team, allineare la remunerazione ai contributi quantificabili di ciascun set di dati garantisce la massima equità nella distribuzione delle royalty. . Indica la sua preferenza per la variante Shapley con dati In- run meno costosa.
	Modellazione della formazione	<b>TRAK: attribuzione del comportamento del modello su larga scala</b> Sung Min Park, Kristian Georgiev, Andrew Ilyas, Guillaume Leclerc, Aleksander Madry MIT	3 aprile 2023	Aleksander Madry, professore al MIT, è responsabile del ragionamento basato sull'intelligenza artificiale presso OpenAI. La soluzione Trak si basa su una rappresentazione approssimativa del modello di intelligenza artificiale per descriverne l'evoluzione durante l'addestramento e mettere in relazione le previsioni con i dati di addestramento che le hanno influenzate.  Questa soluzione si è dimostrata utile con classificatori di immagini addestrati su ImageNet, modelli visione-linguaggio (CLIP) e modelli linguistici (BERT e mT5).

ATTIVITÀ COMMERCIALE	METODO	TEAM DI RICERCA	DATA	MOTIVAZIONE
ADOBE Università Carnegie Mellon Università della California a Berkeley	Somiglianze dei set di dati	<b>Valutazione dell'attribuzione dei dati per i modelli testo-immagine</b> Sheng-Yu Wang, Alexei Efros , Jun-Yan Zhu, Richard Zhan	8 agosto 2023	Poiché è probabile che il metodo della similarità riveli somiglianze casuali, questa distorsione può essere attenuata estendendo il metodo a due serie di immagini, quella di addestramento e quella generata. I ricercatori di Adobe calcolano quindi il contributo del primo set, il dataset , a un set di contenuti sintetici
	Filigrana	<b>ProMark : filigrana di diffusione proattiva per attribuzione causale</b> Vishal Asnani , John Colomosse e altri	14 marzo 2024	Il tatuaggio dimostra la causalità tra contenuto formativo e contenuto sintetico. Ciò consente di basare con sicurezza l'attribuzione legale o i pagamenti ai creatori.
	disimparare	<b>Attribuzione dei dati per modelli testo-immagine mediante il disapprendimento delle immagini sintetizzate</b> Sheng-Yu Wang, Aaron Hertzmann , Alexei A. Efros , Jun-Yan Zhu, Richard Zhang	Giugno 2024	Questo metodo disimpara l'immagine sintetizzata, addestrando le immagini dimenticate per procura, identificando quelle con deviazioni di perdita significative dopo il processo di disimparamento ed etichettandole come influenti.
Marchio Università della British Columbia Università Simon Fraser KTH Istituto Reale di Tecnologia	Shapley federato	<b>Migliorare l'equità per la valutazione dei dati nell'apprendimento federato orizzontale</b> Zhenan Fan, Huang Fang, Zirui Zhou, Jian Pei, Michael P. Friedlander, Changxin Liu, Yong Zhang	23 maggio 2022	La motivazione dei proprietari dei dati dipende in parte dall'equità della collaborazione e dalla ricompensa nell'apprendimento federato.

## Varianti e sviluppi

In tutti i metodi menzionati:

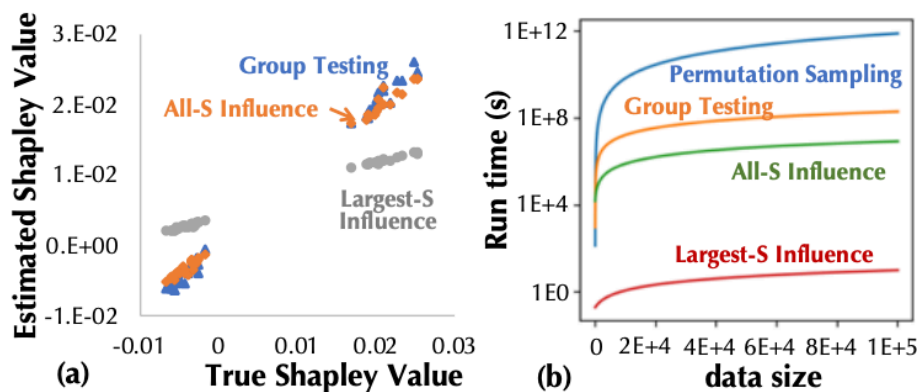
- I parametri di remunerazione sono regolabili in base ad arbitrati sugli indicatori di utilità: dimensione, data, somiglianza, numero di metadati, successo commerciale indipendente, ecc.

Indicatori di utilità

RICERCATORE, AZIENDA	MEDIA	INDICATORI DI UTILITÀ
Piao , Chen e altri	Immagine	Vettori
Wang, Deng e altri	Enciclopedia, Pubblicazione accademica, Immagine, Codice	Modello controfattuale (S) Contenuti generati dal modello $x(\text{gen})$ e dal modello controfattuale
Wang, Mittal e altri	Enciclopedia, Pubblicazione accademica, Immagine, Codice, E-mail, Forum	Mantenere il set di punti dati di convalida $Z(\text{val})$
Deng, Zhang e altri	Musica	Velocità media degli eventi Altezza media di tutti gli eventi Somma delle differenze di tempo di tutti gli eventi
Shutterstock + Bria	Immagine, immagine vettoriale	Punteggio di similarità Ricavi per set di dati Volume dei dati per set di dati e per autore Volume dei metadati Ruolo svolto nello sviluppo dei modelli originali Ruolo svolto attraverso pagamenti di royalty collegati a future attività di licenza generativa
Adobe Stock + Bria	Immagine, immagine vettoriale	Punteggio di similarità Ammissibilità dell'immagine Data dell'immagine Numero di download

- l'algoritmo è regolabile tra precisione e velocità di esecuzione. Il valore di Shapley, in particolare, può essere calcolato
  - con un numero arbitrario di contenuti generati
  - con granularità arbitraria (per set di dati , sottoinsiemi di dati , fino ai singoli dati)
  - con funzioni simili

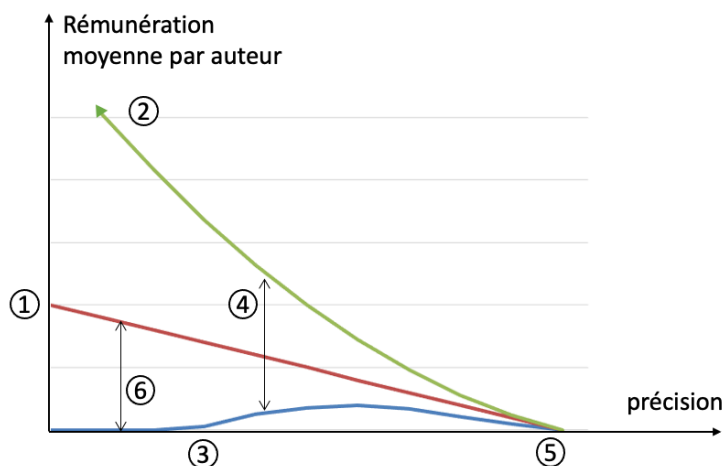
Tre funzioni che approssimano il valore di Shapley



Allo stato attuale dello stato dell'arte, il campo di indagine è sufficientemente ampio per sperare di trovare soluzioni di remunerazione, per utilizzo, che ottimizzino il rapporto tra equità, costo e precisione.

Impatto del costo della precisione sulla remunerazione.

- (19) Con una precisione pari a zero, la remunerazione per autore è pari all'incasso totale diviso per il numero di autori.
- (20) A precisione zero, la massima remunerazione potenziale di un autore è pari all'incasso totale
- (21) A precisione zero, la minima potenziale remunerazione di un autore è pari a zero
- (22) Lacuna di incertezza
- (23) Quando il costo della precisione supera la raccolta totale, il divario di compensazione e incertezza è pari a zero
- (24) A riscossione costante, la remunerazione media per autore diminuisce con precisione



## 7 – Accettabilità

**JDN**  
JOURNAL DU NET

**#FairlyShare contre le péché originel d'Internet** – septembre 2014, par Vincent Lorphelin, Gilles Babinet et al.

Ce qu'ils dénoncent ? Les plate-formes qui reposent sur l'industrialisation du travail gratuit, en dehors des règles de l'appel d'offre, du concours ou du bénévolat. Ils emboîtent ainsi le pas à [...] la levée de boucliers qui a suivi le rachat du journal de crowdsourcing Huffington Post par AOL pour 315 millions de dollars. [Lire](#)

metodo “ pay -to-train” è attualmente il più diffuso per remunerare gli autori le cui opere vengono utilizzate per addestrare modelli di intelligenza artificiale di uso generale. Nonostante la parziale trasparenza e correttezza, non ha ancora dato luogo ad alcuna causa o petizione, come si è visto, ad esempio, in occasione della sensibilizzazione sul "lavoro gratuito" o della vendita dell'Huffington Post (vedi box).

Banche mediatiche che utilizzano modelli “ pay -to-train” e AI utilizzando i loro set di dati

DATA	MEDIA	FORNITORE DI DATASET	FORNITORE DI MODELLI AI
25/10/22	Immagine, foto	Immagine	IA aperta
01/12/23	Immagine, foto	Immagine	Meta , Google , Amazon , Apple
03/12/23	Immagine, foto	Immagine	LG ( Esaone )
21/03/23	Immagine, foto	Immagini Getty	Nvidia
29/05/23	L'Avatar	Arte Ascendente	
07/11/23	Immagine, foto	Immagine	IA aperta
08/08/23	Immagine, foto	Immagine	Nvidia (Picasso)
09/07/23	Immagine, foto	Immagini Getty Alamy Bria	

DATA	MEDIA	FORNITORE DI DATASET	FORNITORE DI MODELLI AI
13/09/23	Musica	AudioSparx	Stabilità AI
13/09/23	Immagine, foto	Adobe Stock (Firefly)	
26/09/23	Immagine, foto	Getty Images (Getty Gen AI)	
10/04/23	Immagine, foto	Canva (Magic Studio)	AI aperta (Dall-e) Google (Immagine)
21/02/24	Immagine, foto	Superstock k Invaso	
04/11/24	Immagine, foto	OcchioEm Scelta gratuita Secchiello fotografico	
27/06/24	Musica, Voce	Musica universale* Intrattenimento Sony Dischi Warner	Google ( YouTube )
30/07/24	3D	Immagine	Nvidia (Edifice)

\*sperimentazione

I contratti di acquisto di set di dati da parte delle banche mediatiche sono stati accompagnati da discorsi che promuovono la remunerazione “rispettosa, responsabile, sicura ed equa” degli autori.

Discorso sull'equità del rapporto retributivo -formativo

PIATTAFO RMA	STOCK E DATASET	DICHIARAZIONI
Nvidia	Getty, Shutterstock, Adobe	Nvidia : "Abbiamo creato una piattaforma che consente ai nostri partner di formarsi partendo da dati concessi in licenza, ad esempio da Getty, Shutterstock o Adobe", ha affermato Huang. "Rispettano i proprietari dei contenuti. I dati di formazione provengono da questa fonte e qualsiasi beneficio economico ne derivi può essere restituito ai creatori. »
	Adobe	Adobe: stiamo sviluppando l'intelligenza artificiale generativa in modo responsabile, con i creatori al centro . Adobe: crea con sicurezza, sapendo che il tuo output sarà della massima qualità, realizzato con modelli progettati per essere sicuri per l'uso commerciale.

PIATTAFO RMA	STOCK E DATASET	DICHIARAZIONI
	Getty	Getty: Generative AI di iStock è uno strumento di intelligenza artificiale generativa conveniente e commercialmente sicuro [...] senza il timore che qualcosa di protetto legalmente si sia insinuato nel set di dati e possa finire nel loro lavoro
Meta	Immagine	Shutterstock : competenza nella creazione di un ecosistema su larga scala che premia e mette in contatto i collaboratori con i creatori.
LG Exaone	Immagine	Shutterstock : La proprietà dei contenuti generati dall'intelligenza artificiale non può essere attribuita a un singolo individuo, ma deve invece essere compensata dai numerosi artisti che hanno partecipato alla creazione di ogni nuovo contenuto. I nostri clienti possono acquistare in licenza le immagini AI per le proprie esigenze, in modo sicuro e legale, senza preoccupazioni. Compensiamo anche i collaboratori i cui lavori addestrano i nostri modelli, così anche tu puoi utilizzare la nostra intelligenza artificiale con la coscienza pulita. Il nostro generatore di immagini AI rimane la funzionalità sicura, facile da usare e progettata in modo responsabile che conosci e ami
		LG: Il modello di intelligenza artificiale è addestrato con milioni di immagini ad alta risoluzione e metadati di Shutterstock e convertirà i prompt di testo in immagini. I collaboratori i cui lavori sono stati utilizzati per addestrare il modello saranno compensati tramite il Contributor Fund di Shutterstock e saranno inoltre compensati ogni volta che i clienti creeranno e concedono in licenza nuovi contenuti generativi che utilizzano la loro proprietà intellettuale.
AI aperta	Immagine	OpenAI : I dati che abbiamo ottenuto in licenza da Shutterstock sono stati essenziali per l'addestramento di DALL-E. Garantire che l'economia dei creativi continui a essere dinamica è una delle massime priorità di OpenAI . Scrittori, artisti, compositori e altri creatori hanno contribuito in modo incommensurabile alle società nel corso della storia della civiltà [...] OpenAI non vuole sostituire i creatori .
	Tela	Canva : il convertitore di testo in immagini di Magic Media, DALL·E di OpenAI e Imagen di Google Cloud [...] garantiscono che i loro modelli di intelligenza artificiale siano equi ed etici. "Condividiamo il valore che sviluppiamo in Canva con i creatori che sono stati con noi per tutti questi anni. Grazie al nuovo Creator Fund , hanno accesso a un pool di royalty equivalente alle creazioni dell'intelligenza artificiale. »
Google Cloud	Tela Immagine	Google Cloud: Shutterstock è diventata anche un'azienda leader nell'innovazione, essendo stata la prima a lanciare un generatore di immagini AI basato su fonti etiche, ora potenziato con Imagen su Vertex AI.
Google Youtube	UMG	YouTube, sussidiaria di Google, sta cercando di stringere accordi con le major musicali per utilizzare l'intelligenza artificiale. Vuole garantire che " i titolari dei diritti musicali vengano compensati per il loro contributo ai dati di formazione. UMG e YouTube sono in trattativa per ottenere la licenza per utilizzare le voci e le melodie degli artisti per addestrare modelli di intelligenza artificiale.
Amazon SageMaker	Alamy Getty Invaso	Alamy : al centro dell'etica del programma Fair Diffusion c'è l'impegno a remunerare equamente fotografi, artisti e creatori. Getty: Aderiamo ai più elevati standard etici e rispettiamo i diritti di proprietà intellettuale e di privacy altrui.
Sony		Il presidente della Sony ha affermato: "I nuovi prodotti e le nuove attività create a partire dalla musica devono essere sviluppati con il consenso del proprietario e con un'adeguata remunerazione e riconoscimento. È fondamentale capire perché si esegue l'addestramento dei modelli di intelligenza artificiale, quali prodotti verranno sviluppati come risultato e qual è il modello di business che consentirà di monetizzare l'utilizzo del lavoro dell'artista.  "Se un modello di intelligenza artificiale generativa viene addestrato sulla musica allo scopo di creare nuove opere musicali che competano sul mercato musicale, allora l'addestramento non costituisce un uso corretto. In questo caso, la formazione non può aver luogo senza il consenso, il riconoscimento e il compenso degli artisti e dei titolari dei diritti. »

## 8 – Brevetti

Nel preparare questo rapporto, abbiamo identificato i seguenti brevetti:

Brevetti relativi alla remunerazione dei lavori di formazione sull'intelligenza artificiale generativa

TITOLO	RIF.	PAESE DEL DEPOSITANTE	CONSEGNATO	PIONIERE
Sistemi e metodi di lavoro creativo sistemi collaborativi	US9100444	NOI	✓	✓
Valutazione del capitale degli stakeholder nei progetti collaborativi	Stati Uniti10776889	NOI	✓	✓
Attribuzione del contenuto visivo generato agli esempi di formazione	Stati Uniti12033372	NOI	✓	
Attribuzione basata su modello per i contenuti generati dall'intelligenza artificiale	US12013891	NOI	✓	
Generazione di vettori autore	US10599770	NOI	✓	
Attribuzione basata sull'input per i contenuti generati dall'intelligenza artificiale	Stati Uniti2024419949	NOI		
Attribuzione basata sull'output per i contenuti generati dall'intelligenza artificiale	US2024419720	NOI		
Sistema e metodo di distribuzione delle commissioni	US2023401550	NOI		
Generazione, utilizzo e attribuzione dei contenuti	Numero di telefono	NOI		
Facilitare la generazione e l'utilizzo di contenuti visivi	Numero di telefono: WO2023084512	NOI		
Metodo e sistema di selezione dei clienti online basato sull'apprendimento federale, supporto e attrezzatura	Numero di parte: CN118446287	Cina	✓	
Generazione dell'indice delle immagini in base alle somiglianze delle caratteristiche dell'immagine	Codice CN103207879	Cina	✓	
Metodo e sistema per valutare il grado di contributo delle caratteristiche locali dell'immagine alle caratteristiche complessive	Numero di parte: CN111652239	Cina	✓	
Metodo e sistema per quantificare i contributi dei clienti nell'apprendimento federato	Numero di parte: WO2023036184	Cina		
Metodo di elaborazione dei dati di distribuzione dei dispositivi mobili di apprendimento federale basato sul valore Shapley	Codice CN114912626	Cina		
Metodo di eccitazione dell'apprendimento federale basato sul valore di cooperazione Sharpley	Codice CN115965092	Cina		
Metodi e sistemi per l'apprendimento federato con predittori locali	Numero di telefono	Cina		
Metodo e sistema di collaborazione	Stati Uniti 9723059	Unione Europea	✓	✓
Metodo per un autoencoder spiegabile e una rete avversaria generativa spiegabile	US11948083	Unione Europea	✓	
Architettura per l'apprendimento rinforzato spiegabile	Stati Uniti11455576	Unione Europea	✓	
Tracciabilità dei contributi multifase nell'ambiente di creazione collettiva	Stati Uniti2015363743	Unione Europea		
Metodo e dispositivo per premiare la creazione collaborativa	EP2958068	Unione Europea		



Questa lista dimostra che, quando si tratta di remunerare i titolari dei diritti, gli europei non sono superati da Cina e Stati Uniti, il che contrasta con la loro posizione di outsider per quanto riguarda l'intelligenza artificiale in generale e l'IA generativa in particolare.

### The EU's position in complex (digital and green) technologies

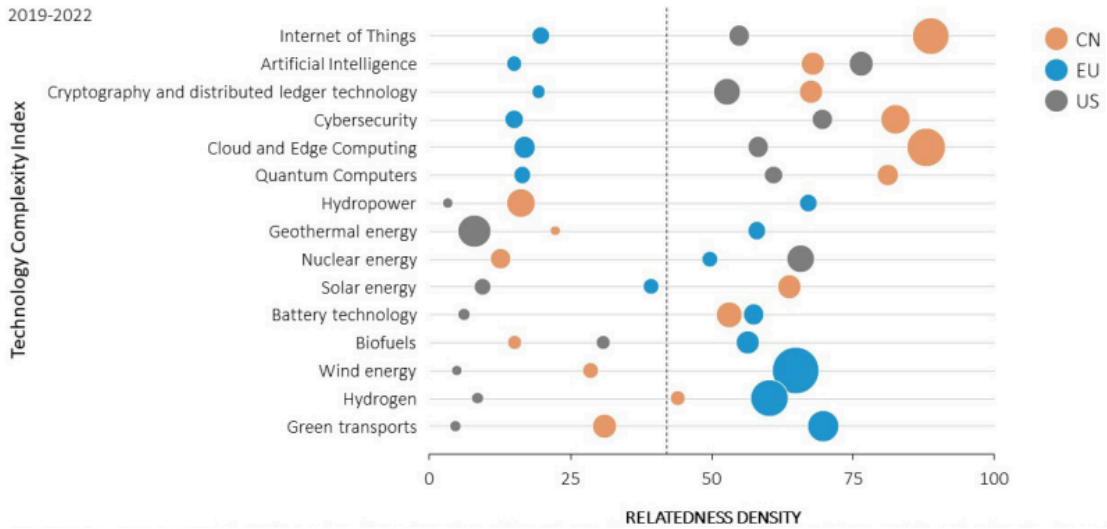


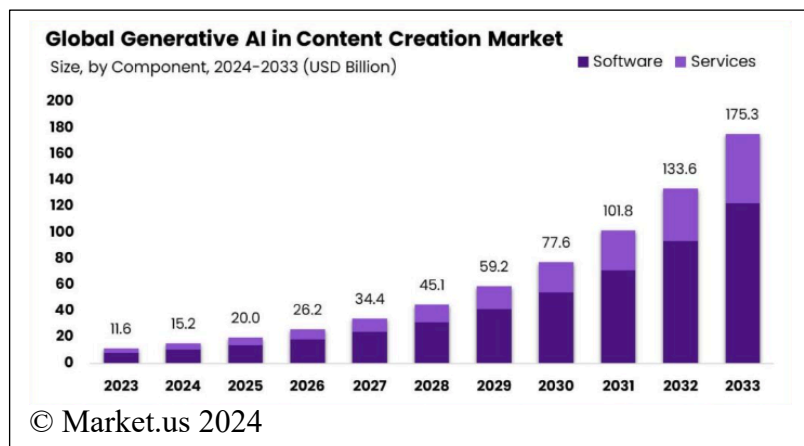
Table 7 Top patent owners in GenAI modes (companies), 2014–2023

	Images/ video	Text	Speech/ voice/ music	3D image models	Molecules/ genes/ proteins	Software/code	Other modes
Tencent Holdings (China)	607	565	551	102	57	41	464
Ping An Insurance (China)	262	600	599	26	33	24	223
Baidu (China)	395	465	441	81	26	18	166
IBM (US)	101	274	168	13	37	42	132
Alibaba Group (China)	142	213	144	36	2	10	133
Samsung Elect. (Republic of Korea)	173	140	226	34	10	5	57
Alphabet/Google (US)	138	107	200	25	15	15	72
ByteDance (China)	173	82	112	21	6	7	90
Microsoft (US)	78	194	151	18	6	22	42
BBK Electronics (China)	125	58	117	12	0	4	116
Netease (China)	78	76	80	37	0	11	102
NTT (Japan)	37	100	169	11	0	3	62
Huawei (China)	104	84	72	24	1	4	102
China Mobile (China)	79	67	83	10	0	7	89
State Grid (China)	75	54	31	6	1	2	144
Adobe (US)	190	79	34	30	0	9	20
Sony Group (Japan)	87	35	105	42	0	1	42
Siemens (Germany)	124	25	8	10	3	4	59
Ant Group (China)	45	82	29	3	0	6	59
Industrial and Commercial Bank of China (China)	40	62	49	5	0	16	48

Source: WIPO, based on patent data from EconSight/IFI Claims, April 2024.

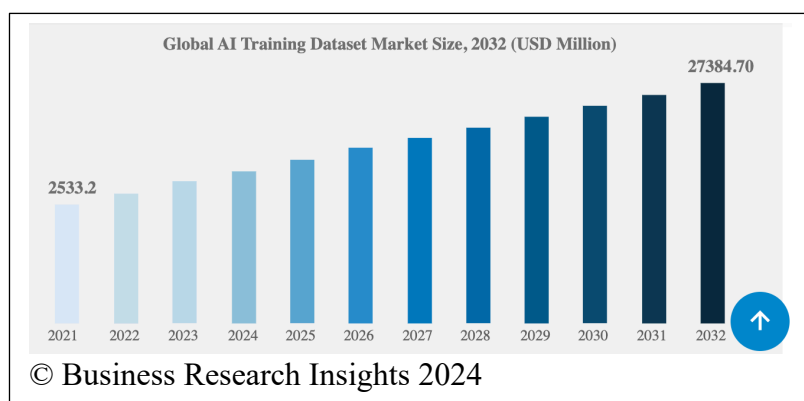
**Prospettive**

## 9 – Mercato



Solo per il segmento della creatività, si stima che il mercato globale dell'intelligenza artificiale generativa raggiungerà gli 11,6 miliardi di dollari nel 2023 e supererà i 100 miliardi di dollari entro il 2031.

Per dare un'idea della portata, un tasso di remunerazione equo del 15% rappresenterebbe più dell'attuale riscossione delle OGC (Organizzazioni di gestione collettiva) per i diritti artistici (musica, audiovisivi, spettacoli dal vivo, letteratura e arti visive).



Le aziende specializzate in set di dati di addestramento all'intelligenza artificiale organizzano la raccolta, l'annotazione, la cura dei dati e il loro adattamento alle esigenze degli algoritmi di apprendimento.

il mercato globale dei set di dati di addestramento all'intelligenza artificiale valeva 2,5 miliardi di dollari e si prevede che raggiungerà i 27 miliardi di dollari entro il 2032.

### Modelli di compensazione per gli autori di dati di intelligenza artificiale generativa

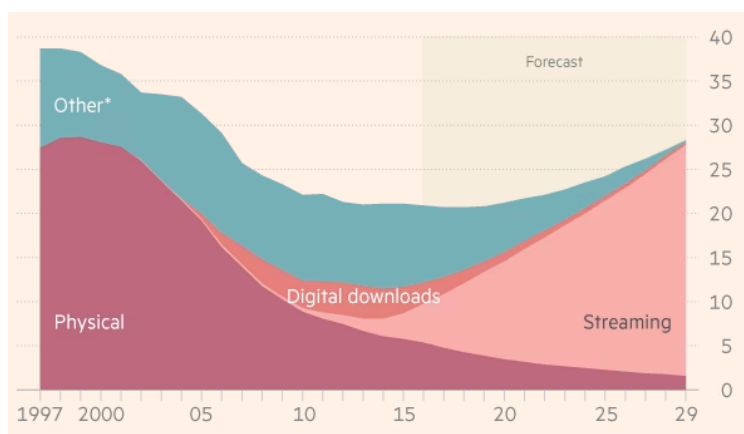
MODELLO DI BUSINESS	AZIENDE	COMPENSO DELL'AUTORE
Raschiatura (1)	OpenAI , Midjourney , Stabilità, MistralAI	NO
Uber dei dati (2)	Mturk , ClickWorker , Gigwalk , Premise	Pagamento all'ordine
Banche di contenuti (3 )	Shutterstock, Getty, Canva, Adobe	Pagare per allenarsi
Gruppi mediatici (4)	Axel Springer, Financial Times, Stack Overflow, Reddit, Le Monde	NO
Aziende degli autori (5)	la Corporazione degli Autori , Creato da Human, Human Native AI, Abbastanza addestrato, Dataset Providers Alliance	Retribuzione equa

- Scraping : DALL·E, ad esempio, è stato inizialmente addestrato con centinaia di milioni di immagini didascaliche provenienti da Internet. Questa pratica, che è la maggioranza, è causa di numerose cause legali.
- Uber di dati: i creatori di modelli di intelligenza artificiale hanno iniziato a coprire i rischi e a proteggere le catene di fornitura dei dati, sia attraverso un'industria fiorente di broker di dati sia attraverso accordi con i proprietari di contenuti emersi per soddisfare la domanda. Sta emergendo un settore di aziende specializzate in dati di intelligenza artificiale, che creano reti di lavoratori con contratti a breve termine per produrre contenuti personalizzati su richiesta. I prezzi variano da 1 a 2 dollari a fotogramma, da 2 a 4 dollari per un video breve e da 100 a 300 dollari all'ora per i filmati più lunghi. La tariffa di mercato per un testo è di \$ 0,001 a parola. Ai proprietari delle foto, dei podcast e dei dati medici viene pagato circa il 20-30% dell'importo totale della transazione.
- Banche di contenuti: Shutterstock , Getty, Canva o Adobe dispongono di grandi basi multimediali fornite dagli autori, pagati in base ai set di dati venduti. Shutterstock , leader nelle immagini, ha venduto licenze ai venditori di intelligenza artificiale per oltre 100 milioni di dollari, ovvero il 12% del suo fatturato totale, e si prevede che crescerà del 40% nel 2024. Queste licenze di immagini, video e musica formeranno un mercato indirizzabile di 10 miliardi di dollari nel 2030, con una crescita annua del 22% nello stesso periodo.
- Gruppi di media: i proprietari di database concedono in licenza i set di dati , senza compensi aggiuntivi <sup>50</sup>gli autori
- Società di autori: gli autori vogliono poter definire le loro opzioni di licenza per l'intelligenza artificiale generativa. Stanno fiorendo idee come quella dell'*Authors' Guild* con la piattaforma *Created by Human* , come *Human Native AI* , *Fairly Formati* , *Dataset Providers Alliance* o *AI Commission* in Francia. In questa fase, i modelli economici non sono stati definiti e le prospettive di sviluppo non sono state quantificate.

## Proposte di licenza per opere di intelligenza artificiale generativa

INIZIATIVA	PRINCIPALI PROPOSTE
La Gilda degli Autori	<ul style="list-style-type: none"> <li>- Stipula di accordi per l'utilizzo delle opere</li> <li>- Pagamento dei diritti d'autore per l'utilizzo delle opere protette</li> <li>- Trasparenza sui dati utilizzati</li> </ul>
Creato da Umano	<ul style="list-style-type: none"> <li>- Etichetta “ Creato dall’Uomo ” per le opere</li> <li>- Piattaforma che consente agli autori di autorizzare e controllare le licenze</li> </ul>
Intelligenza artificiale nativa umana	<ul style="list-style-type: none"> <li>- Mercato per set di dati di formazione di intelligenza artificiale generativa</li> <li>- Autorizzazioni di tracciamento per ogni contenuto utilizzato</li> </ul>
Abbastanza addestrato	<ul style="list-style-type: none"> <li>- Etichetta “ Abbastanza Addestrato » per l'intelligenza artificiale generativa</li> <li>- Contratti standardizzati per l'utilizzo di opere in ambito AI</li> </ul>
Alleanza dei fornitori di set di dati	<ul style="list-style-type: none"> <li>- Remunerazione equa</li> <li>- Contratti standardizzati ( pagamento per formazione, royalties sul fatturato, abbonamento, licenza per settore)</li> <li>- Trasparenza sull'uso dei dati di formazione</li> <li>- Autorizzazioni individualizzate per autore</li> </ul>
Commissione AI	<ul style="list-style-type: none"> <li>- Piattaforma pubblica</li> <li>- Intermediazione della Biblioteca Nazionale di Francia e dell'Istituto Nazionale Audiovisivo</li> </ul>

Facendo un parallelo con il mercato musicale, ci troviamo attualmente nella fase “ Napster -IA”, in attesa di soluzioni equivalenti a Spotify e Apple Music. I loro modelli di business hanno ripulito il mercato a partire dal 2012, quando streaming e pirateria venivano precedentemente confusi.



© Financial Times, Goldman Sachs

mercato, verrà liberato dalla proposta di nuovi modelli economici del tipo delle "società d'autore" e dalla loro esplorazione volontaria da parte degli stakeholder.

Nonostante la loro natura embrionale, la moltiplicazione di queste iniziative apre una nuova prospettiva di “proprietà liquida”.

Allo stesso modo, crediamo che il mercato potenziale per la valorizzazione delle opere tramite l'intelligenza artificiale generativa sia maggiore di quello degli OGC e rappresenti un'opportunità per i titolari dei diritti. Simmetricamente, gli operatori di intelligenza artificiale generativa potrebbero accedere a basi di contenuti di qualità.

Questo potenziale, che per il momento avvantaggia segmenti alternativi del

## 10 – Verso la proprietà liquida

### Interview de Vincent Lorphelin

Le Soir (Bruxelles), Janvier 2024



"Sam Altman ed Elon Musk afferma che l'intelligenza artificiale generativa renderà il lavoro obsoleto , quindi preparatevi al reddito universale. Ciò si coagula con la tesi della fine del lavoro diffusa da Jeremy Rifkin qualche anno fa . Ciò si collega anche a quello della singolarità che suggerisce che entro pochi anni , l'intelligenza artificiale generale andate superano gli esseri umani in tutti i campi. La prova è che abbiamo sempre detto che l'intelligenza artificiale non entrerà mai nelle professioni. creativa (non diventerà mai una scrittrice, una designer...). Solo che dall'anno scorso siamo lì. Le porte del bastione sono sfondate . [...]

Al contrario, il DNA europeo è l' economia sociale di mercato , è l' economia che crea la società . Il lavoro ha un senso. Storicamente e da un punto di vista strategico , la gestione collettiva dei diritti è due volte più sviluppata

in Europa che negli Stati Uniti . Pertanto, se torniamo a questa questione fondamentale di interesse generale , per controbilanciare la tesi della singolarità , l' interesse generale L'Europa deve promuovere il modello decentralizza .

Immaginiamo che questo modello di piattaforma decentralizzata diventi sistemico , diventi un sistema , allo stesso modo in cui lo fanno oggi le piattaforme centralizzate come Gafam . In questo caso la questione manipolati , si tratta di diritti intellettuali, non di merci. Questo è un valore economico non di mercato . L' economista Karl Polanyi sosteneva che l' economia di mercato aveva mercificato il lavoro, mentre il lavoro non può essere equiparato a una merce.

Le piattaforme centralizzate cercheranno il più possibile di dare un valore di mercato ai diritti intellettuali. In una logica decentralizzata , i diritti saranno espressi in percentuali, non in valuta. Valori economici manipolati non sono più espressi in euro o dollari. È un'economia che non si basa sul denaro, ma sulla proprietà . Per tracciare un parallelo con ciò che Emile Zola chiamava "denaro liquido" (con il passaggio da un'economia in cui la ricchezza non proveniva più dalla proprietà terriera ma dalla circolazione del denaro), si propone la nozione di " proprietà liquida ", dove il capitale circolante dell'economia è la circolazione dei diritti di proprietà intellettuale .

Oggi abbiamo finanziarizzato tutto a tal punto che un'azienda presenta i suoi conti come se fosse in vendita ogni giorno , perché misura tutto in denaro. Esistono però attività immateriali , fuori bilancio, che esprimono valori non espressi in denaro. La ricchezza dell'azienda va oltre ciò che esprime il denaro. Quando manipoliamo le proprietà dei liquidi , siamo portati a trovare altri indicatori. "Ci stiamo muovendo gradualmente verso una società che rende solvibile la ricchezza di un altro ordine". (leggi di più [qui](#)).

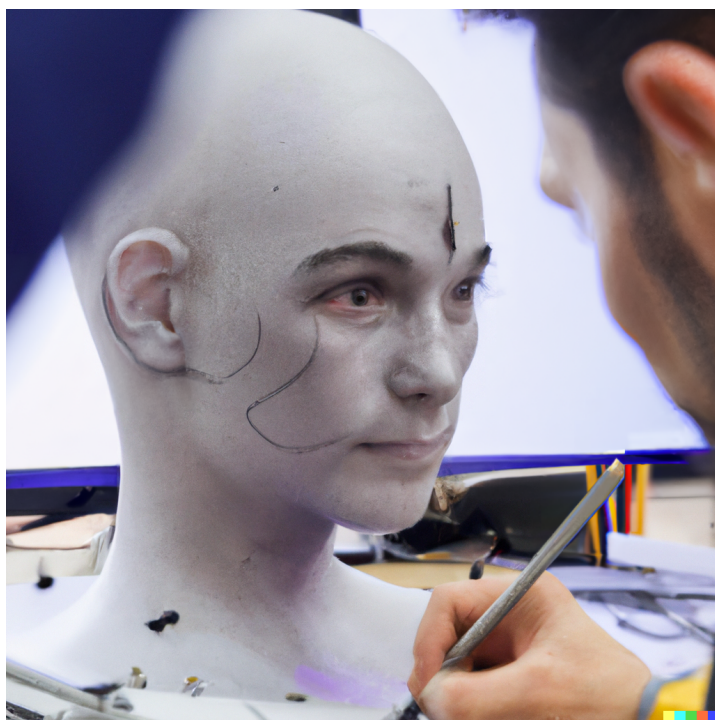
Così come il modello di piattaforma di Uber ha ispirato il concetto di uberizzazione, chiamiamo "sacémisation" il fenomeno della diffusione del modello economico delle società del diritto d'autore, che valorizzerà il capitale umano attraverso la proprietà liquida.

## 11 – Promuovere l'intelligenza umana

Brad Smith, presidente di Microsoft, sta commettendo un errore morale nei confronti dell'azienda. Parlando dell'intelligenza artificiale generativa come ChatGPT, afferma: "Abbiamo tutti il diritto, in base alla legge sul copyright, di leggere e imparare. Ora ci chiediamo se possiamo permettere alle macchine di imparare allo stesso modo. maniera. Penso che ci sia un imperativo sociale per rendere ciò possibile. »

Questo intervento può sembrare banale. Dopotutto, se le macchine assorbono Internet per creare articoli e illustrazioni a prezzi ridicolmente bassi, perché privarcene noi stessi? Il New York Times e gli altri titolari

di diritti che si oppongono a questo utilizzo dei loro contenuti non stanno forse dando una brutta lezione a Microsoft? Se le macchine non commettono plagio, che male c'è? Questo è il risultato finale inaccettabile di questo discorso.



### Una vantaggiosa vaghezza giuridica

le piattaforme Gafam raccolgono una moltitudine di piccoli contributi e promuovono il lavoro gratuito. Cambiano ancora dimensione con l'intelligenza artificiale, la cui voracità richiede la conquista di nuovi territori, quelli degli autori.

In via eccezionale, il legislatore aveva definito il "fair use" autorizzando Google a pubblicare estratti di contenuti privati. Sebbene l'esito del processo sia ancora incerto, l'ex avvocato Brad Smith sa che l'opinione pubblica avrà un ruolo importante

. Ecco perché egli distilla l'immagine mentale adulterata del "diritto delle macchine ad apprendere". Il suo "imperativo " " sociale " non è che una volgare contestazione della legittimità degli autori a difendere i propri territori.

Distrae dibattendo su false questioni etiche. L'intelligenza umana è destinata a essere declassata? Possiamo lasciare che sia l'intelligenza artificiale a prendere decisioni? Potrebbe benissimo adottare la postura di un uomo saggio di fronte a un Elon Musk è meno cortese, ma entrambi sono presagi della stessa illusione: quella di superare l'intelligenza umana.



## **IA: un'illusione di superiorità ?**

Il miglior romanzo non sostituirà mai l'esperienza di vita . La migliore “ videochiamata ” non sostituirà mai un incontro reale. Per quanto sofisticato possa essere il modello , non è quello reale . Ognuno di noi ha un'intelligenza del mondo che supererà sempre quella delle macchine, le quali non hanno altra intelligenza se non quella dei loro modelli .

Le IA illuminano la nostra intelligenza. Permettono alla polizia ambientale di individuare gli inquinatori, all'insegnante di comprendere le difficoltà di mille studenti o al laboratorio di anticipare gli effetti di un farmaco .

Diffondere l'illusione di un'imminente superiorità delle macchine è irresponsabile. Si tratta di negare l'opera propriamente umana, che trae ispirazione proprio dal mondo e dai suoi modelli , quando l'artista rivela qualcosa di non detto, quando il conducente compensa le leve che rispondono male , quando il medico dubita di una diagnosi troppo ovvia , o quando il giudice ha una convinzione intima.

### **La difesa degli autori**

Alimentare l'illusione di una competizione tra intelligenze e attribuire diritti umani alle macchine è una colpa nei confronti della società . Al contrario, dobbiamo affermare la preminenza dell'intelligenza umana e garantire l' integrità dei suoi territori.

L'illusione di Brad Smith non avrebbe alcun senso se non fosse per la figura più importante al mondo nel campo dell'intelligenza artificiale. Conquistata da questa illusione, la Commissione interministeriale sull'intelligenza artificiale afferma che le macchine supereranno gli esseri umani. Lei difende , fortunatamente, la giusta remunerazione della creazione umana , ma senza rendersi conto che è la valorizzazione anche di questa accuratezza che è manipolata .

L'Europa ha già deciso con l'AI ACT di proteggere il copyright. Ora deve istituire il catasto dei territori sfruttati dalle IA e difenderne la rendita. ( [Per saperne](#) di più leggi Les Echos del 6 maggio 2024)



## 12 – Una soluzione di sacemizzazione per l'Europa

### Entretien avec Jean-Paul Betbèze

Economiste, Membre du Comité scientifique de la Fondation Robert Schuman



### Quali dovrebbero essere i punti chiave della “ sacemizzazione ”?

soluzione “ sacemizzante ” per l'Europa dovrebbe:

- Basarsi sulla proprietà intellettuale degli autori, come gli OGC. È necessario preservare e garantire la proprietà intellettuale degli autori, che altrimenti potrebbe dissolversi o scomparire nell'attuale rivoluzione della comunicazione. La comunicazione sta diventando globale, è alla base del grande cambiamento che stiamo vivendo, con la moltiplicazione dei messaggi in tutte le direzioni. Il contributo di ciascuno deve essere riconosciuto, a meno che non si voglia inaridire l'attuale movimento mondiale con un'eccessiva concentrazione a vantaggio di alcuni autori, mentre gli altri non troveranno alcun vantaggio nel loro lavoro e nelle loro idee, nessun riconoscimento, nessuna remunerazione. A causa delle lacune nella

disponibilità di capitale, in gran parte a vantaggio degli Stati Uniti , il pericolo è una monopolizzazione delle apparenti fonti di creazione.

- Raccogliere royalties sullo sfruttamento delle opere da parte dell'IA. La “ sacemizzazione ” non solo ci permette di rintracciare, ma anche di identificare e stimare le fonti, anche quando sembrano minime. Questo processo consente di preservare tutte le origini e quindi di riconoscere tutte le creazioni e i miglioramenti: nulla si perde, tutto si conserva.
- Riconoscere la creatività e distinguerla dai contenuti amatoriali. Nella proliferazione delle opere è importante saper mettere in risalto quelle originali, valorizzarle e promuoverle.
- Attuare meccanismi di incentivazione individuale, affinché il prezzo della creatività sia competitivo sul mercato e compatibile con i costi di produzione. Il rischio di moltiplicare il lavoro è la banalizzazione, soprattutto perché il costo fisico di produzione è ovviamente basso. Una fatturazione legata esclusivamente al costo di produzione rischia di essere disincentivante . Il grande vantaggio della “ sacemizzazione ” è quello di allontanarsi da una logica di prezzo legata al costo di produzione, per passare ad un'altra logica, quella dei mercati, che si basa sull'importanza del successo, sulla ripetizione del messaggio.

## **Qual è il vantaggio di un mercato guidato dalla “ sacemizzazione ”?**

In questa logica di mercato, è la domanda misurata, da cui nasce la domanda anticipata, che crea i nuovi prezzi e quindi i margini. In questo contesto i margini non possono che aumentare, quindi gli utili marginali e quindi le valutazioni. Questa differenza tra costi bassi, rilevati dai motori di ricerca, e valutazioni che avranno così una base oggettiva, aggiornate dagli strumenti tipo Sacem , è alla base della logica proposta. Ciò consentirà di valorizzare il contributo di tutti, soprattutto nel contesto europeo.

## **E per le attività creative?**

Non dimentichiamo che la concorrenza monopolistica è sempre all'opera: spinge verso la riduzione del numero dei concorrenti, finché non ne emerge uno solo: il monopolio. Lì, più che altrove, il vincitore si prende tutta la posta in gioco. La “ sacemizzazione ” deve opporsi a ciò consentendo una migliore distribuzione della remunerazione prima che appaia un singolo monopolista. Ciò è ovviamente favorevole alla ricerca, all'innovazione e alla creazione, poiché sappiamo che la situazione di monopolio si rivolta contro il monopolista stesso, in una situazione di mercato che a lungo termine si esaurisce, perché le sue dinamiche portano a un calo dei suoi profitti. La “ sacemizzazione ” rispetta quindi le creazioni e rallenta la selezione spietata e, in ultima analisi, distruttiva del monopolio stesso.

Nel mercato delle idee la velocità è mortale, la memoria dei contributi è decisiva, unita alla loro giusta remunerazione. Oggi è possibile e, comprensibilmente, più necessario che mai. Si tratta di salvare il creato attraverso il progresso offerto dalla tecnologia. »

Michelle Bergadaà, Professeure émérite à l'université de Genève.

Co-autrice de *Réinventer l'intégrité académique à l'ère de l'intelligence artificielle* (à paraître, EMS, février 2025)



### **Lei afferma che la legge sull'intelligenza artificiale non riguarda solo gli autori, ma anche i ricercatori. Perché?**

MB: “Dobbiamo prendere in considerazione l'applicazione dell'AI Act per compensare gli autori di “ricerche autentiche”. La Francia paga male i suoi ricercatori, il che rallenterebbe la loro fuga all'estero o verso il settore privato. Per la società costa molto di più schiacciare i suoi giovani ricercatori sotto la burocrazia che pagarli bene. Molti si demotivano e dedicano ore extra all'insegnamento invece di dedicarsi alla ricerca. Quindi, se lo Stato non è in grado di pagarli adeguatamente, deve farlo il sistema di produzione di articoli, libri e relazioni. Questo deve essere il pilastro della strategia europea per l'intelligenza artificiale generativa. Con il vantaggio aggiuntivo di un sistema automatizzato di rilevamento delle frodi.

### **La frode non è sempre esistita?**

La crisi degli oppioidi negli Stati Uniti [il Fentanyl e farmaci correlati, prescritti su prescrizione medica, da soli sono stati responsabili di circa 71.000 decessi per overdose nel 2021] è nata a causa di un singolo articolo falso, erroneamente citato e ristampato. Questo accadeva molto prima dell'intelligenza artificiale. Grazie all'intelligenza artificiale, la diffusione di articoli fraudolenti avviene

principalmente tramite nuovi articoli. Questo è un nuovo modello di sciame del falso.

Oggi coesistono tre temporalità. In primo luogo, il tempo istantaneo ed emotivo (di Trump e della Generazione Z) consente la produzione non solo di informazioni false, ma anche di falsi autori, altrettanto credibili. In secondo luogo, l'eterno presente delle comunità di una volta preserva un mondo alternativo in cui la Terra è piatta e la candeggina elimina il Covid. Il terzo tempo lineare è guidato dalla tecnologia materialistica di Elon Muschio. L'intelligenza artificiale è un acceleratore di frodi che prospera negli interstizi di queste temporalità.

### **Come agire?**

Pubpeer e Retraction Watch Database cercano di catturare i singoli trasgressori, uno per uno. Ma ci vollero tre anni per completare l'indagine e ottenere una ritrattazione sui giornali.

Questo fenomeno deve essere affrontato sistematicamente con radar di integrità, un'ingiunzione ai truffatori di restituire il denaro ricevuto e l'esclusione Banche dati di intelligenza artificiale. Il futuro della nostra società sono i giovani che educiamo. Aiutiamoli assumendo una posizione ferma e remunerando le loro creazioni originali attraverso l'intelligenza artificiale generativa. Ciò motiverebbe molti giovani ricercatori onesti a creare conoscenza. »

**Raccomandazioni**

## 13 – La soluzione Controv3rse

Sam Altman, fondatore di Open AI, vuole avere "nuovi modelli di business che consentano ai creatori di generare entrate. »

Abbiamo visto che l'Uber del modello di dati emerge per evitare la degenerazione dell'intelligenza artificiale generativa, ma che minaccia il capitale umano. Il modello economico deve basarsi sulle OGC per aprire una prospettiva di rivalutazione di questo capitale attraverso la proprietà liquida e la sacemizzazione .

Il modello di business che consigliamo a Sam Altman e ad altri operatori dell'intelligenza artificiale generativa è il seguente:

### Soluzione Controv3rse

RACCOMANDAZIONE	AI APERTA	OGC
1 - Riserva	Riservare il 15% del fatturato dell'IA generativa per una giusta remunerazione degli autori e dei titolari dei diritti	Fornire agli autori una piattaforma per opporsi, ove opportuno, all'uso delle loro opere da parte dell'intelligenza artificiale generativa, o per scegliere una licenza, opera per opera
2 – Segmentazione	Identificare l'OGC principale per segmento di mercato (media/ dataset * paese)	Rivendica un segmento di mercato (media/ set di dati * paese)
3 - Metodo	Pubblicare il metodo di attribuzione da parte di OGC, dettagliare la disambiguazione in caso di recupero	Negoziare accordi di coordinamento o reciprocità tra OGC
4 – Non attribuzione	Identificare un fondo per paese per gli importi non assegnati	Gestione dei reclami del fondo
5 - Relazioni	Pubblicare il riassunto con sufficienti dettagli  Inviare a ciascun OGC identificato un report mensile contenente gli identificativi degli autori, i dettagli dell'uso delle loro opere, per contenuto sintetico e il valore attribuito	Invia all'operatore l'elenco delle opere che non possono essere utilizzate dall'IA generativa
6 - Pagamento	Distribuire la riserva agli OGC e ai fondi	Distribuire gli importi raccolti agli autori e ai titolari dei diritti

## 14 – Inizia la negoziazione

### Passi

Per raggiungere accordi tra gli operatori di intelligenza artificiale generativa e i CMO, raccomandiamo che le parti interessate documentino i seguenti passaggi:

#### Fasi della soluzione Controv3rse

RACCOMANDAZIONE	PROCESSO	MIGLIORI PRATICHE, COMPARABILI
1 - Prototipo	Discutere a partire da una soluzione tecnica da perfezionare per facilitare la proiezione degli interlocutori in un risultato e ridurre la complessità del dibattito (vedi il nostro dimostratore)	Metodi dimostrativi, di prova ed errore
2 - Vision	Sviluppare la lungimiranza: decentralizzazione vs. piattaforme centralizzate, sacemizzazione vs. Uberizzazione dell'intelligenza artificiale, lavoro degli autori vs. mercificazione, absurdità del "diritto" della macchina, proprietà liquida, metriche di utilità	Quadro europeo: economia sociale di mercato, economia della creatività, patrimonio e diversità culturale
3 - Equa remunerazione	<p>Specificare il giusto tasso di remunerazione basato sulla misurazione dell'utilità di tutti i lavori degli autori per tutti i contenuti generati. Offriamo un valore iniziale del 15%.</p> <p>Proporre indicatori dell'utilità dei contributi per l'attribuzione autore per autore e il metodo per perfezionare tale attribuzione in previsione di futuri progressi tecnici.</p>	<p>Spargere Direttiva 2006/115/CE (13) Diritti connessi (73)</p> <p>Statistiche Yacast per Spré</p>
4 – Codice di buona condotta dell'Ufficio AI	<p>Osservatorio all'avanguardia</p> <ul style="list-style-type: none"> <li>- Metodi: valore Shapley, pay -to-train, similarità</li> <li>- Arbitrati: parametri, granularità, indicatori di utilità</li> <li>- Piattaforme di opzioni e licenze</li> <li>- Identificazione dei titolari dei diritti e degli autori</li> <li>- Elaborazione degli importi non assegnati</li> </ul> <p>Riassunto sufficientemente dettagliato:</p> <ul style="list-style-type: none"> <li>- Contenuto</li> <li>- Granularità</li> <li>- Formato</li> </ul> <p>Segnalazioni ai titolari dei diritti:</p> <ul style="list-style-type: none"> <li>- Rivendicazioni di dataset da parte dei titolari dei diritti</li> <li>- Modalità di allocazione e distribuzione da parte del titolare dei diritti</li> <li>- Elaborazione dei recuperi</li> <li>- Dati</li> </ul> <p>Contenuto segreto: informazioni non divulgate nei riassunti e nei rapporti</p> <p>Normalizzazione progressiva</p>	<p>Codice di buone pratiche contro la disinformazione</p> <p>Raccomandazioni dell'Authors ' Guild</p> <p>Missione del CSPLA</p> <p>ISWC, ISAN</p>
5 – Brevetti	Garantire posizioni francesi ed europee attraverso un portafoglio di brevetti dedicati, in cooperazione pubblico-privato	Programma PI dell'Alleanza INPI

## Retribuzione equa

Un'equa compensazione attraverso l'intelligenza artificiale generativa potrebbe ispirarsi a quella dello Spré . In Francia, questa organizzazione riscuote una royalty calcolata sul fatturato dei locali notturni, a beneficio dei musicisti. Un campione rappresentativo di 120 di essi è dotato di un box “ Yacast ”, finanziato da Sacem e Spré . Yacast raccoglie le playlist della musica trasmessa in streaming e pubblica statistiche per artista, etichetta, genere e numero di settimane di trasmissione. Dopo la ritenuta legale del 9%, lo Spré distribuisce il prelievo agli OGC (SCPA 50%, ADAMI 25% e SPEDIDAM 25%) e comunica loro le statistiche Yacast per remunerare i beneficiari finali (artisti interpreti, produttori, azioni di interesse generale). .

## Tasso equo

Nel suo rapporto del 2020, *L'autore e l'atto della creazione*, Bruno Racine deplorava il fatto che “la remunerazione proporzionale degli autori raggiunge o supera solo eccezionalmente il 10% dei ricavi operativi”. Egli attribuisce questa osservazione al «rapporto che lega l'artista-autore agli attori a valle (editori, distributori, produttori) [che] appare profondamente squilibrato, il che lo porta, per quanto lo riguarda, a mettere in discussione in molti casi l'idea stessa di libertà contrattuale.” Per porre rimedio a ciò, egli ha voluto "rafforzare collettivamente gli artisti-autori, mediante la rapida organizzazione di elezioni professionali che consentano di dare sostanza e legittimità al Consiglio nazionale degli artisti-autori che verrà creato per fungere da cornice per negoziazione collettiva con le emittenti" per stabilire "un sistema contrattuale che retribuisca il tempo di lavoro dell'autore, precluso dallo sponsor".

Poiché l'idea di rivalutazione del valore del lavoro non è applicabile al caso dell'IA generativa, proponiamo di definire il tasso equo come il valore dell'utilità di tutti i contributi degli autori riferiti alla figura aziendale del contenuto generato. Questo è il principio su cui si basa il valore Shapley e l'economista Ernst Fehr stima che il valore apportato dalla stampa a Google Search sia del 14% . Poiché la famosa domanda "ti è stato utile questo commento ? " » lanciato da Amazon, tutti i siti web commerciali migliorano la qualità della loro offerta moltiplicando le metriche di utilità, ormai di uso comune.

Fissare la base imponibile sul fatturato dei contenuti generati consentirebbe di non far dipendere la giusta remunerazione dell'autore dalla maggiore o minore segmentazione verticale del settore, ad esempio: autore / piattaforma / dataset / modello generale / messa a punto / interfaccia / generazione.

## 15 – Licenza – Riconoscimenti

Questa nota e le sue tabelle sono accessibili con licenza CC-BY, subordinatamente alla giusta condivisione dei frutti del suo sfruttamento da parte dell'intelligenza artificiale generativa, concessa dall'associazione Controv3rse, RNA W751267044.

Questo documento è stato tradotto automaticamente da Google Translate. In caso di ambiguità prevale la versione francese.

Trae ispirazione dalle riflessioni e dalle discussioni tra i membri delle missioni ministeriali per la legge sull'intelligenza artificiale , quelle dei think tank indipendenti Controv3rse e Institut de l' Economie , nonché altre personalità riconosciute per la loro competenza. Si tratta di un ampliamento dello studio già realizzato per la missione parlamentare Pradal -Rambaud e delle rubriche pubblicate su *Le Monde* e *Les Echos* .

Vincent Lorphelin, fondatore e co-presidente di questi think tank, autore di questa nota, li ringrazia calorosamente per le loro idee e per la qualità di queste discussioni.

Presidenti e Relatori delle missioni parlamentari e ministeriali:

Alexandra Bensamoun , professoressa di diritto all'Università di Parigi Saclay, ex membro della Commissione per l'intelligenza artificiale

Joëlle Farchy , professoressa universitaria, direttrice di m2ecn e della cattedra PcEn Università Paris 1 Panthéon-Sorbonne, membro del CSPLA

Philippe Pradal , Deputato della terza circoscrizione delle Alpi Marittime, Assemblea nazionale

Stéphane Rambaud, Deputato della terza circoscrizione del Varo, Assemblea nazionale

Lionel Ferreira, Maestro delle Richieste presso il Consiglio di Stato

Julie Groffe -Charrier, docente HDR di diritto privato presso la Facoltà Jean-Monnet

Bastien Blain, professore associato di economia, scienziato dei dati e del comportamento, Università di Parigi 1 Panthéon-Sorbonne

Esperti e imprenditori:

William Bailey  
Co-fondatore e presidente, Bolero Music

Direttore Marketing, Comunicazione e Innovazione, Keyrus

Emmanuele Benazera  
Presidente e CEO, Jolibrain

Alice Coatalem  
Co-fondatore di CogNeed e professore associato, Università Paris Dauphine PSL

Jean-Paul Betbèze  
Economista , Betbèze Conseil

Christophe Collet  
Fondatore, AskLocala

Michelle Bergada  
Presidente dell'Istituto per la ricerca e l'azione contro le frodi accademiche e il plagio

Federico Dayan  
Fondatore, Exactcure

Michel Bokobza  
Responsabile della missione, Collège de Paris

Guillaume Desveaux  
Co-fondatore di Aleia e amministratore, AI Cargo Foundation

Laurent Cervoni  
Direttore della ricerca e dell'innovazione, Talan

Dory di Jonathan  
Fondatore , Live Crew

Charles-Eric de la Chapelle  
Fondatore, Myriad Data

Guglielmo Eldin  
Fondatore, XXII

Jean-Philippe Clair

Stephanie Flacher  
Co-fondatore, Logion Network



Raffaello Frisch  
Co-fondatore, HawAI.tech

Alain Garnier  
Amministratore delegato, Jamespot

Antonio Graveline  
Fondatore, Disaitek

Filippo Guguen  
Presidente, Sorga - Mappa Emulsione

Francesco Hachem  
Fondatore, Codenekt

Rodolphe Hasselvander  
Fondatore, Blue Frog Robotica

Brice Hoarau  
Fondatore, Semdee

Matthias Houllier  
Co-fondatore, Wintics

Casey Joly  
Avvocato, IPso

Jean Latger  
Fondatore, Oktal -SE

Frédéric Lefebvre- Naré  
Direttore dei dati AI, Niji

Giovinezza Lemrabet  
Fondatore, Everysens

Jacques Levy- Vehel  
Presidente di Case Law Analytics e  
Direttore della ricerca, INRIA

Olivier Laborde  
Responsabile dell'innovazione e della  
trasformazione digitale, BPCE

Alessandro Leforestier  
Fondatore, Panodyyssey

Sistina Lorphelin  
Ingegnere di intelligenza artificiale, UTC,  
studente presso l'INSEAD

Aymerico Masurella  
Co-fondatore, Spoon.ai

Pietro Miralles  
Co-fondatore, Footovision

Emanuel Moyrand  
Co-fondatore, Francia Meta

Clemente Merville  
Fondatore, Manzalab - Teemew

Edoardo di Miollis  
Fondatore, Polycube

Nathalie Nevejan  
Professore di diritto privato e intelligenza  
artificiale, Università di Artois

Remy Ozcan  
Co-fondatore, Crypto4All

Jean-Jacques Quisquater  
Professore di crittografia, Polytechnique  
Louvain, editorialista di Le Monde

Jean-Michel Salomon  
Presidente, Società degli autori di giochi

Federico Soufflet  
Co-fondatore, Haapie

Tequi Clemente  
Co-fondatore, Terno

Francois-Xavier Thoorens  
Amministratore delegato, Vaultys

Arnaud Touati  
Avvocato associato, #Hashtag


Christophe Tricot  
Co-fondatore, La Forge

Killian Vermersch  
Co-fondatore e CEO, Golem.ai

# La solución Controv3rse

Un modelo económico para mejorar el capital cultural y humano a través de la IA generativa



 Adobe Firefly

CONTROV3RSE

# Resumen

Controv3rse es un think tank independiente, formado por 70 emprendedores y expertos en IA generativa. Este informe ofrece una solución en respuesta a la declaración de Sam Altman, fundador de OpenAI, quien aboga por un nuevo modelo económico que recompense a los creadores ante el auge de la IA generativa.

Hay que decir que el conflicto entre los autores, cuyas obras están siendo saqueadas, y los explotadores de la IA generativa, que necesitan claridad y estabilidad jurídica, no se resolverá en los tribunales hasta dentro de al menos 10 años.

Mientras tanto, la proliferación de la IA generativa diluye la originalidad y borra los rastros de las obras fuente. Al mismo tiempo, las plataformas de microtarefas están respondiendo a la demanda masiva de datos en detrimento de modelos basados en la creatividad y los derechos de autor, como SACEM. El capital cultural y humano, antaño central para la estrategia europea, se ve doblemente amenazado por su extracción.

Sin embargo, el economista y premio Nobel Jean Tirole destaca que la riqueza nacional se basa en la promoción de la innovación, apoyada en la propiedad intelectual como palanca de la creatividad. Para preservar el capital cultural y humano y estimular la innovación, el nuevo modelo económico debe reinventar el de las sociedades de autores.

Lloyd Shapley introdujo en 1953 un concepto clave para distribuir equitativamente las ganancias en un juego cooperativo, basado en la contribución de cada jugador. El método derivado, ideal para remunerar a los autores de IA generativa, lamentablemente tiene un costo computacional prohibitivo. Existen alternativas, como el modelo de “pago por formación” utilizado por los bancos de imágenes (Shutterstock, Getty). Compensan a los autores en función de la cantidad de contenido proporcionado, pero ignoran la calidad o utilidad de las obras. Las empresas emergentes y los investigadores están complementando este modelo para ofrecer una distribución más justa basada en la similitud. Ofrecemos al lector un simulador online que le permitirá comprobar por sí mismo, a escala real, la viabilidad técnica y económica de este enfoque.

En perspectiva, este modelo protege los derechos de los autores y se opone a la “uberización de la IA”. La “sacrificación” valora la “propiedad líquida”, la riqueza intelectual de la creatividad, que mantiene y desarrolla el capital cultural y humano frente a su extracción por los monopolios.

Por ello, recomendamos a Sam Altman y a los operadores de IA generativa un modelo económico en cooperación con las sociedades de autores, y a todas las partes interesadas un método para iniciar negociaciones en este nuevo marco de reflexión.

**Desafíos**

# 1 – ¿Quién quiere que la IA generativa pague a los autores?

"Creo que necesitamos un nuevo acuerdo, un nuevo estándar, un nuevo protocolo, como quiera llamarlo, sobre cómo se recompensará a los creadores", dijo Sam Altman, fundador de Open AI. "Necesitamos tener nuevos modelos económicos que permitan a los creadores generar ingresos. »

La idea de pagar a los autores a través de gen-AI ha sido defendida por otras personalidades durante varios meses.

## Operadores de IA que quieren pagar a los autores

NOMBRE	COMPAÑÍA	CITA
Sam Altman	IA abierta	"Creo que necesitamos un nuevo acuerdo, un nuevo estándar, un nuevo protocolo, como quiera llamarlo, para determinar cómo se recompensará a los creadores. Necesitamos nuevos modelos económicos que permitan a los creadores generar ingresos. »
Arthur Mensch	Inteligencia artificial Mistral	"Tenemos interés en tener acceso a contenidos de calidad. Por ello, hemos iniciado conversaciones con proveedores de contenidos, tanto del sector editorial como de la prensa. Se pueden encontrar sinergias y mecanismos para compartir valor. »
David Holz	A mitad de camino	"No hay forma de tomar cien millones de imágenes y averiguar de dónde provienen. Sería bueno que las imágenes tuvieran algunos metadatos incorporados sobre el propietario de los derechos de autor o algo así. Pero esto no es posible, no hay registro. No hay forma de encontrar una imagen en Internet y luego rastrearla automáticamente hasta su propietario y luego poder hacer algo para autenticarla. »
Chris Dixon	a16z	"El movimiento [Web 3] tiene el potencial de recuperar el espíritu de los primeros tiempos de Internet, proteger los derechos de propiedad de los creadores, recuperar la propiedad y el control de los usuarios y romper el dominio de las grandes empresas centralizadas en nuestras vidas. "
Neal Stephenson	Lamina1 (inventor del metaverso)	"En realidad, puedes rastrear cuánta influencia tuvo una imagen o entrada determinada en el resultado final. Y luego, si ese resultado final vale algo, entonces tal vez haya una manera de dar crédito y una compensación proporcional por esos aportes [...] Digamos que estuviera escribiendo un libro que tenía una espada mágica dentro que simplemente estaba descrita en el libro. [...] Alguien podría entonces crear un activo que pudiera venderse en Unreal Asset Store [...] Hay diseñadores de sonido que tienen que hacer algo similar con los sonidos que produce el juego cuando se utiliza. Y tienes programadores que tienen que, usando planos o C++, integrar la espada en el juego para que realmente sea capaz de hacer algo y contribuir a la experiencia. Así que en cada paso se añade más valor. Y al final de ese proceso, tienes algo que realmente podría generar ingresos. Y cuando finalmente lleguen esos ingresos, lo que nos gustaría que hicieran es propagarse hacia atrás. ¿Y le gustaría que las diferentes personas que contribuyeron a la cadena de valor fueran compensadas de alguna manera?

NOMBRE	COMPAÑÍA	CITA
Jaron Lanier	Microsoft	“A las personas se les pagará por sus datos y también pagarán por servicios que requieran datos de otras personas. La atención de los individuos estará guiada por sus intereses autodefinidos en lugar de por plataformas manipuladoras en deuda con anunciantes u otros terceros. Las plataformas recibirán datos de mayor calidad con los que entrenar sus sistemas de aprendizaje automático. » Lanier reconoce que incluso los investigadores que estudian la dignidad de los datos no pueden ponerse de acuerdo sobre cómo desenredar todo lo que los modelos de IA han absorbido o cuánto detalle se necesita para tenerlo en cuenta. Sin embargo, Lanier cree que es posible hacerlo, de manera gradual. “Necesitamos calcular y presentar la procedencia de las fuentes humanas que fueron más importantes para un resultado de IA determinado. Actualmente no hacemos eso. Podemos hacerlo de manera efectiva y eficiente, simplemente no lo hacemos. “Debe ser una decisión social dar este paso”.
Yat Siu	Marcas de Animoca	“Los datos son el recurso más valioso. Para empezar, impulsan toda la inteligencia artificial de la que tanto oímos hablar. Sin datos, no hay ChatGPT . Los datos son el nuevo trabajo. Y no recibimos una compensación justa por ello. [...] Existe una clara correlación entre los derechos de propiedad y la riqueza de las naciones. » “Imagínese entrar a Facebook y al final del día darse cuenta de que acaba de ganar mil dólares para la empresa. Tu relación con Facebook cambiaría por completo. Probablemente exigirías tu parte justa. »
Yann Le Cun	Meta	"un modelo de propagación de los derechos [de autor] en una dirección y de los ingresos en la otra"
Viaje de Adler	Ex director ejecutivo de Scribd	“La Cuarta Ley es un conjunto de principios rectores sobre cómo las empresas de IA pueden usar y entrenarse con contenido generado por humanos. La cuarta ley, inspirada en las Tres leyes de los robots del autor de ciencia ficción Isaac Asimov, establece que los humanos deberían tener derecho a consentir y controlar cómo la IA utiliza su trabajo, a recibir una compensación (si se solicita) y a recibir crédito por su trabajo (si se trata de un libro). se hace referencia en el resultado, debería haber un enlace para comprarlo). »

Dado el liderazgo de Sam Altman, la idea de compensar a los autores adquiere más sustancia, aunque considera que esta compensación solo sería debida cuando se plagia el estilo del autor, lo cual es muy limitante.

Sin embargo, cuestiona el modelo económico de su implementación . El objetivo de esta nota es proponer este nuevo modelo económico en el marco de la Ley de IA .

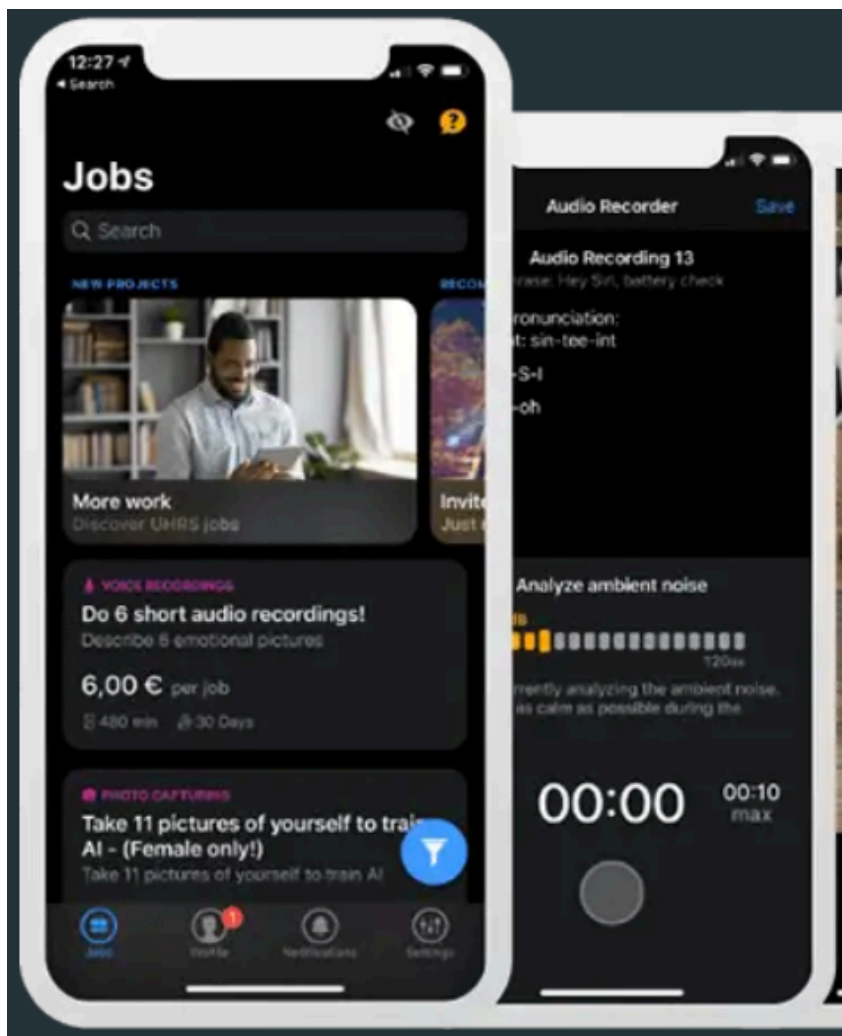
## 2 – Las tres caras de la remuneración de los autores

Si se aborda a un nivel puramente legal, el conflicto sobre el scraping , que enfrenta a los autores con los operadores de IA generativa, durará al menos diez años.

Mientras tanto, la industria se segmentará en creadores de datos, plataformas, agregadores, negociadores, formadores, refinadores, explotadores e integradores de IA generativa, lo que diluirá el acto de creación. El patrimonio cultural se habrá mezclado, difundido y reciclado en contenidos cuyos rastros originales se habrán borrado.



Por temor a una degeneración de los contenidos de aprendizaje, ya se está desarrollando un nuevo sector de microtareas para la producción en masa de contenidos.



©Clickworker 2024

de economía colaborativa para IA se basan en la lógica de comandos. Clickworker paga a los usuarios 6€ directamente por describir 6 imágenes en voz alta y publicar la grabación. O tomarte una foto haciendo ejercicios físicos. Gigwalk y Premise pagan \$4 para tomar una fotografía de un atasco de tráfico en una intersección. El resultado se utiliza para entrenar IA.

Este modelo tipo “Uber de datos” está creciendo rápidamente gracias a la enorme necesidad de datos por parte de los operadores de IA generativa.

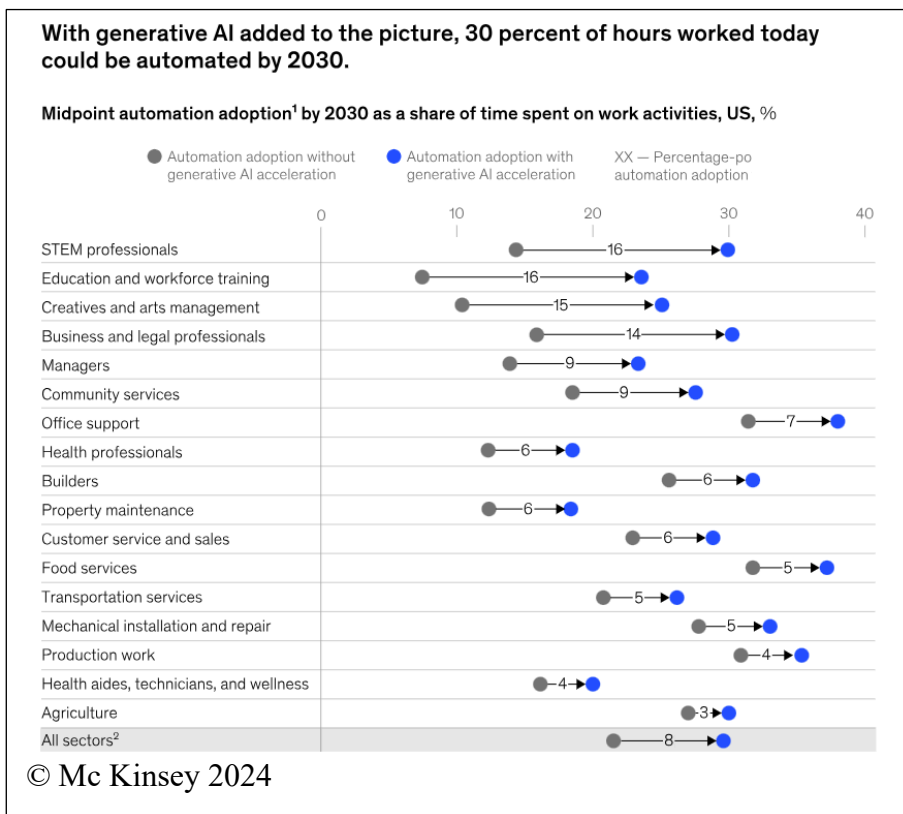
Se trata de lo opuesto a las sociedades de derecho de autor como la SACEM, basadas en la oferta y la propiedad intelectual de los autores, que rastrean esta propiedad, recaudan regalías y remuneran a los titulares de derechos según escalas de distribución predefinidas.

El primer modelo favorece la demanda, la ejecución, la mercantilización inmediata, el neotaylorismo cognitivo y la estandarización del trabajo. El segundo promueve la oferta, la iniciativa, la propiedad intelectual, la creatividad y la diversidad cultural. El primero explota el capital cultural, el segundo lo cultiva.

Entre otros modelos económicos (ver más abajo), el scraping, el data Uber y las sociedades de autores son las tres caras principales de la remuneración —o no— de los autores, cuyo desequilibrio agrava las tensiones actuales.



### 3 – Más allá de los derechos de autor, el capital humano



Frente al desarrollo de la IA generativa, los autores y la creatividad están en primera línea. A continuación vendrán la artesanía, las profesiones artísticas, la especialización, la investigación y todas las profesiones que implican saber hacer, en todos los sectores económicos.

A medida que el modelo del “Uber de los datos” se generalice, dará lugar a una forma de Uberización de la IA, de la misma manera que la empresa Uber inició en su día el fenómeno de la Uberización.

Por ejemplo, personalizar la IA generativa empresarial con todos sus archivos de datos permite obtener un borrador de

cada nuevo mensaje, informe, proyecto, presentación o precio. Las correcciones profesionales entrenan a la IA hasta que la tasa de error se vuelve aceptable. Las ganancias de productividad son significativas pero, a cambio, las posiciones de la empresa en los eslabones estratégicos de creación de valor se reducen en beneficio del operador de IA.

Como todas estas competencias son la intersección de conocimientos, prácticas y materiales de organización social, lo que está en juego es el tejido de las empresas. Es como si, en la época de la revolución industrial, un cártel de propietarios de máquinas de vapor controlara la organización taylorista del trabajo. Esto es lo que el mercado de valores anticipa para las grandes tecnológicas. Al conceder capitalizaciones de billones de dólares, ya no valora las empresas, sino los monopolios.

Se difundirá la idea de que las máquinas tienen “derecho a aprender” gratuitamente. Dentro de diez años, incluso si los autores acaban ganando su batalla legal contra el scraping, su compensación se basará en los comparables proporcionados por los Ubers de datos, es decir, no mucho. En retrospectiva, nos daremos cuenta de que los derechos de autor existentes no eran más que una Línea Maginot que era fácil de eludir.

La economía cuenta con un capital cultural y humano, identificado desde hace tiempo como un activo considerable, hasta el punto de haber sido situado en el centro de la estrategia económica europea. Esta estrategia, que fue demasiado temprana, finalmente quedó olvidada por falta de un modelo económico. Subvaluado y mal defendido, este capital ahora está amenazado de ser extraído por la IA uberizante. Para cualquier empresa, país o continente, rechazar la tecnología significa condenarse a una decadencia duradera, como la de China desde hace dos siglos. Aceptarlo sin compensación es dejarse colonizar sin luchar.

El Premio Nobel Jean Tirole nos enseña que 1/ la riqueza de las naciones depende de su capacidad de capturar valor a nivel de innovación. 2/ La innovación surge de la iniciativa individual, porque la característica del trabajo creativo es que no sabemos qué vamos a encontrar. 3/ La propiedad intelectual es un mal necesario para estimular esta creatividad. 4/ Impedir que una empresa comercialice sus productos hasta que haya pagado sus regalías es un arma muy eficaz contra los Gafam . 5/ Los pools de propiedad intelectual con autorización de licencias individuales requerirían que adquirieran licencias, pero a un precio competitivo.

Las naciones que quieran mejorar su capital humano y cultural deben encontrar una solución sacrificial moderna , cuyo modelo económico pueda competir eficazmente con la IA uberizante .

**Solución técnica**

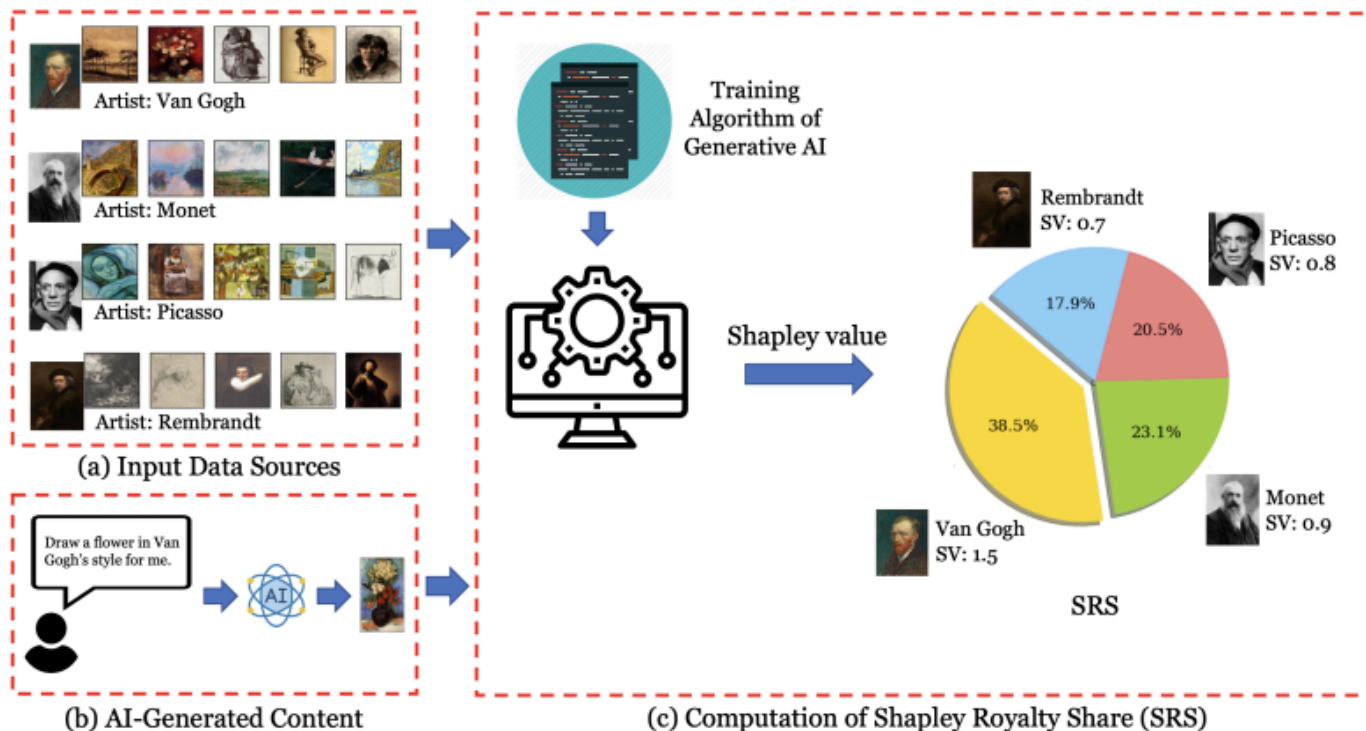
## 4 - Solución ideal

En teoría de juegos, Lloyd Shapley introdujo en 1953 la idea de que era posible, en un juego cooperativo, calcular una distribución justa de ganancias entre los jugadores. Por ejemplo, consideramos un ecosistema marino con todas sus especies vivas (algas, anémonas, conchas, estrellas de mar) y medimos la utilidad de cada especie por su contribución a la vida en el ecosistema. Cuando se eliminan las estrellas de mar, desaparece el 50% de las otras especies, frente al 10% que desaparece cuando se eliminan las anémonas. El valor Shapley de las estrellas de mar es cinco veces mayor que el de las anémonas.

Existe un consenso entre los investigadores sobre la solución ideal para remunerar a los autores. Esto implica calcular el valor Shapley para cada autor y distribuir la remuneración proporcional a este valor.

En el caso de la IA generativa, la distribución justa de la remuneración de los trabajos de formación debería realizarse idealmente en función de la utilidad de los trabajos de formación para el contenido generado por la IA, por autor. En la siguiente figura, el valor Shapley de Van Gogh es 1,5 o el 38,5% de la compensación total.

Valor de Shapley



En la práctica, el coste de calcular el valor Shapley es prohibitivo, lo que plantea la cuestión de la viabilidad técnica de una solución de remuneración justa.

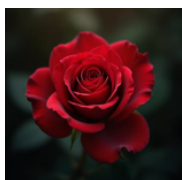
## 5 – Demostrador

### Objeto

El lector encontrará un simulador en línea en <https://huggingface.co/spaces/TheFrenchDemos/genai-rewa rd>







Creemos que ahora es técnica y económicamente posible calcular una puntuación de atribución para el contenido de formación con el fin de distribuir equitativamente la remuneración entre los titulares de derechos de este contenido. Este simulador permitirá a cada uno verificar por sí mismo sus propias hipótesis económicas.

Por ejemplo, se generó una imagen de una rosa con el mensaje " *Una rosa roja vibrante en plena floración, fotografía macro* ":



El simulador identifica en pocos segundos, entre los diez millones de contenidos de la base de datos Open Images, las ocho imágenes originales que más han contribuido a la formación del contenido generado. Calcula una tasa de atribución, compensación y proporciona información de identificación sobre el contenido original y su autor.

#### Atribuciones de imágenes de rosas

CONTENIDO	ATRIBUCIÓN	RECOMPENSA DE CONTENIDO (c€)	URL DEL CONTENIDO	NOMBRE DEL AUTOR	URL DEL AUTOR
	21%	0,63	<a href="https://www.flickr.com/photos/50093642@N03/4629227563">https://www.flickr.com/photos/50093642@N03/4629227563</a>	Yoko Nekonomanía	<a href="https://www.flickr.com/photos/nekonomania/">https://www.flickr.com/photos/nekonomania/</a>
	18%	0,54	<a href="https://www.flickr.com/photos/22305657@N02/11733854773tt">https://www.flickr.com/photos/22305657@N02/11733854773tt</a>	Bruno Pinho	<a href="https://www.flickr.com/photos/brunopinho/">https://www.flickr.com/photos/brunopinho/</a>
	15%	0,45	<a href="https://www.flickr.com/photos/12692992@N00/170278141">https://www.flickr.com/photos/12692992@N00/170278141</a>	Decano y Barb	<a href="https://www.flickr.com/photos/barbanddean/">https://www.flickr.com/photos/barbanddean/</a>
	12%	0,36	<a href="https://www.flickr.com/photos/33140999@N07/3509399376">https://www.flickr.com/photos/33140999@N07/3509399376</a>	Michael Arrighi	<a href="https://www.flickr.com/photos/arrighi/">https://www.flickr.com/photos/arrighi/</a>
	10%	0,30	<a href="https://www.flickr.com/photos/55658968@N00/8730223835">https://www.flickr.com/photos/55658968@N00/8730223835</a>	Toshihiro Gamo	<a href="https://www.flickr.com/photos/dakiny/">https://www.flickr.com/photos/dakiny/</a>
	9%	0,27	<a href="https://www.flickr.com/photos/34226407@N04/8734050294">https://www.flickr.com/photos/34226407@N04/8734050294</a>	Príncipe Edd	<a href="https://www.flickr.com/photos/edp-pics/">https://www.flickr.com/photos/edp-pics/</a>

	9%	0,27	<a href="https://www.flickr.com/photos/41353201@N07/4267426012">https://www.flickr.com/photos/41353201@N07/4267426012</a>	El músico Loxton	<a href="https://www.flickr.com/photos/dj-dwayne/">https://www.flickr.com/photos/dj-dwayne/</a>
	8%	0,24	<a href="https://www.flickr.com/photos/64574638@N04/6006728032">https://www.flickr.com/photos/64574638@N04/6006728032</a>	serbiastruhpro	<a href="https://www.flickr.com/photos/serbestruhpro/">https://www.flickr.com/photos/serbestruhpro/</a>

En la parte inferior de la página web, tres botones permiten al usuario descargar:

- Este informe
- Informe de los titulares de derechos que el operador podría generar y comunicar a los titulares de derechos en aplicación de los acuerdos de liquidación y retrocesión de remuneraciones equitativas
- El informe reglamentario previsto por la Ley de AI (art. 53, 1, d), según el modelo propuesto por el Consejo Superior de las Artes y las Letras Propiedad (CSPLA)

## Cómo funciona

El simulador tiene un botón "avanzado" parámetros » que permite al usuario ajustar los parámetros que desee. La siguiente tabla copia los que están configurados por defecto e indica las fórmulas de cálculo.

### Parámetros y valores de una simulación

VARIABLE	CONFIGURACIÓN	COMPARABLE	CÁLCULO	FÓRMULA
Suscripción de usuario	12 €/mes (1)	Adobe Firefly desde \$4,99/mes Midjourney desde \$10/mes Dall-e incluido con ChatGPT Plus \$20 Getty Edify 45€ por 25 generaciones		S
Contenido generado	60 /mes	de Adobe Firefly 100/mes Midjourney 100/mes		GRAMO
Valor nominal por contenido generado			0,2 €	$V = S / G$
Porcentaje de titulares de derechos y autores				$R\% = Rr\% + Ra\%$
- Titulares de derechos	10%	Música (ASCAP, BMI, SACEM) 2 – 8% Imagen de stock 55 – 85%		% de RR
- Autores	5%	Libros impresos 5 – 15% Música (intérprete y compositor) 15 – 30%		Real academia de bellas artes%

## Fotografía de archivo 15 – 45%

Remuneración de los titulares de derechos y autores por contenido generado	3 céntimos	$R = V \times R\%$
Atribución	21%	$A\% = \text{Algoritmo de similitud}$ $N = \text{número de contenidos asignados}$
Recompensa por contenido de capacitación	0,63 c€	$T = R \times \%$
Número de autores	1.800	B
Usuarios	500.000	Se estima que el número de suscriptores de pago de Midjourney está entre 2 y 5 millones . tú
Recompensa mensual para el titular de los derechos y el autor	829 €	$S \times U \times \% R / B \times \% A \times N$
- Titulares de derechos	276 €	$S \times U \times \% Rr / B \times \% A \times N$
- Autor	553 €	$S \times U \times \% Ra / B \times \% A \times N$

## Concepto de semejanza

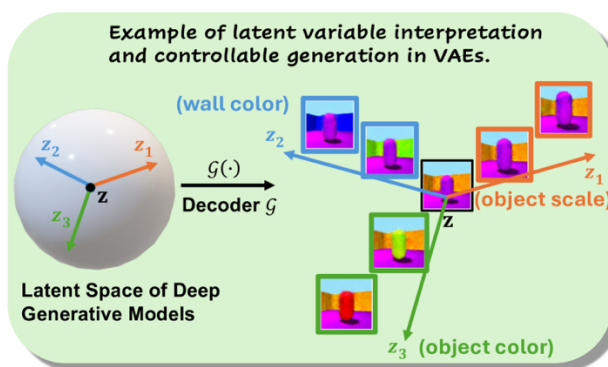
Este demostrador utiliza un algoritmo de similitud entre el contenido de entrenamiento y el contenido generado, implementando [DINOv2](#) y [FAISS](#). Intuitivamente, dos imágenes similares parecen iguales.

Sin embargo, la "similitud" de la que aquí se habla es una noción científica separada de cualquier noción jurídica de "falsificación", "semejanza" o violación de una obra, definida en el derecho francés como "toda reproducción, representación o distribución, por cualquier medio". "medio cualquiera que sea, de una obra intelectual en violación de los derechos de autor, tal como están definidos y regulados por la ley" (CPI, art. L. 335-3). No existe necesariamente infracción o violación por la reproducción de un elemento de una obra, una "contribución" en este caso. De la misma manera, cuando un autor crea, un elemento puede entrar en su composición, sin que medie falsificación o infracción de tercero. La remuneración resultante de la "similitud" no debería por tanto paralizar la acción del autor en caso de infracción.

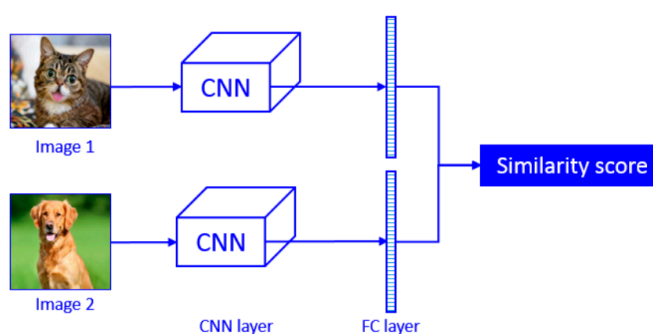
## Similitud vectorial

Técnicamente, cada contenido generado u original se analiza para extraer información medible: gato, perro, color y tamaño del primer plano, tipo, estilo, tamaño, objetos reconocidos, palabras clave, ... todos estos cientos de indicadores de utilidad, alrededor de 500 en nuestro demostrador, forma un vector en un "espacio latente".





El grado de atribución de una obra (imagen 1) a un contenido generado (imagen 2) se calcula mediante una puntuación de similitud entre las dos imágenes, procesada por las dos redes neuronales siamesas. Cuanto más cercanos sean los vectores resultantes, mayor será la puntuación de similitud entre las imágenes.



## Número de autores

La base de datos Open Images se formó a partir de la de Flickr, aplicando filtros como la no detección por parte de Google Image Search, un número mínimo de píxeles o la deduplicación. Open Images también selecciona únicamente licencias CC-BY, es decir, una imagen de cada 30. Dado que Flickr ofrece la licencia "todos los derechos reservados" de forma predeterminada, el cambio de licencia solo lo realizan autores informados. Esto sugiere una alta concentración en Open Images de los colaboradores más activos, habiendo publicado más de 5.623 fotografías, o alrededor de 1.800 autores. Esta hipótesis intuitiva se pudo verificar con un recuento exacto de los autores de la base de datos Open Images, lo que no pudimos lograr dado el límite en la velocidad de consulta.

## Identificación de los autores

La base de datos Open Images identifica a los autores de las imágenes y proporciona su página de perfil de Flickr, que, sin embargo, no incluye una dirección de correo electrónico o postal completa.

Más allá de nuestro demostrador y de Open Images, los conjuntos de datos públicos proporcionan elementos de identificación, de precisión variable, pero que pueden cruzarse para identificar a cada titular de derechos.

Los métodos de referencias cruzadas son más conocidos por su uso fraudulento contra la privacidad. Ya en 1997, la comunicación obligatoria de estadísticas agregadas por parte de una aseguradora de Massachusetts, junto con los datos electorales, permitió identificar sin error posible al gobernador William Weld y conocer su historial médico.

Su uso legal permite, en la mayoría de los casos, la identificación completa de los autores: por comparación de gráficos sociales, por correlación estadística gracias a una rara combinación de atributos (nicho científico, horas de actividad, dirección IP), por reconstrucción de atributos individuales. a partir de promedios grupales y atributos conocidos de otros individuos, etc. La siguiente tabla proporciona pistas para identificar a los autores del contenido de capacitación de GPT 4.



AI-generated by Adobe Firefly

Posible identificación de los autores de los conjuntos de datos utilizados por GPT-4

	CONJUNTO DE DATOS	ELEMENTOS DE IDENTIFICACIÓN	CRUCE DIRECTO	IDENTIFICADOR
<b>Identificación directa</b>	Imágenes abiertas	-	-	Apellido, nombre, ficha de perfil
	Proyecto Gutenberg	-	-	Apellido, nombre, fechas de nacimiento y muerte
	arXiv	-	-	Nombre, Inicial del primer nombre
	PubMed	-	-	Id_autor
<b>Identificación indirecta</b>	Rastreo común	Dominio	Quién es	Registrant_contact (nombre, organización, calle, ciudad, estado, teléfono)
	Corpus de libros	Texto sin formato	Palabras aplastantes	Nombre, primer nombre
	Texto web	Texto simple + búsqueda = "escrito por"	API de Amazon	Nombre, primer nombre
	Desbordamiento de pila	HTML	Stackoverflow.com	ID de usuario , perfil de red
<b>En esperando reclamo</b>	Wikipedia	Nombre de usuario	-	-
	Reddit	Nombre de usuario	-	-
<b>Otro</b>	Voz común	(estadísticas)	-	-
	Ngramas de Google Books	( no aplica )	( n / A )	( n / A )
	GitHub	( no accesible )		

## 6 – Soluciones existentes

### Síntesis

La siguiente tabla identifica algunos métodos comercializados o probados, con respecto a nuestro demostrador:

#### Métodos para promocionar el contenido del autor

VALUACIÓN	MÉTODO DE VALUACIÓN	DEMOSTRADOR	EMPRESAS COMERCIALES	INVESTIGADORES
Contenido generado	Coste de suscripción/cantidad	Controversia		
Conjunto de datos	Similitud de conjuntos de contenidos En forma de shapley Datos de ejecución de Shapley Frecuencia de acceso al sitio web, tipo de datos		peaje	Adobe IA abierta IA abierta
Trabajo de formación	Similitud de contenido Modelado del aprendizaje Personalización de plantillas Desaprender Tatuaje Shapley federado	Controversia	Bria (imágenes), Prorata Gist.ai (textos)	IA abierta Adobe Adobe Adobe Huawei
Autor	Pagar para entrenar  Influencia de la obra en su conjunto Identificación de los titulares de derechos		Shutterstock , Getty , Audiosparx , Canva  Sureel.ai (música)	
		Controversia		

## Soluciones comercializadas

### Empresas que pagan a los autores por contenidos de formación

COMPAÑÍA	CAPITAL DE RIESGO	VOLUMEN DE NEGOCIOS	MÉTODO	MEDIOS DE COMUNICACIÓN	PROVEEDOR DE CONTENIDO
Adobe (luciérnaga)	-	21 mil millones de dólares	Pagar para entrenar	Imagen	
Lienzo	-	2 mil millones de dólares	Pagar para entrenar	Imagen	
Getty	-	917 millones de dólares	Pagar para entrenar	Imagen	
Shutterstock	-	875 millones de dólares	Pagar para entrenar	Imagen	
Prorrato (Gist.ai)	25 millones de dólares	-	Similitud de contenido	Texto	Sky News, Time, Financial Times, The Atlantic, Axel Springer, The Guardian, dmg media, Universal Music, Fortune
Bria	24 millones de dólares	-	Similitud de contenido	Imagen	Shutterstock, Getty Images, Alamy , Superstock
peaje	24 millones de dólares	-	Frecuencia de acceso al sitio web	Texto	Penske Media Corporation, Time, Mumsnet, marcas de medios de confianza, Candr Media Group y AdWeek
Sureel.ai	-	-	Influencia de la obra en su conjunto	Imagen, música	
Audiosparx	-	-	Pagar para entrenar	Música	

## Método de pago por formación

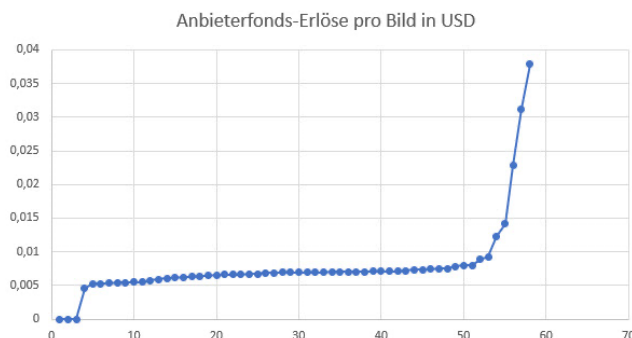


AI-generated by DALL-e

Los bancos de contenido privados de Adobe, Canva, Getty y Shutterstock, distribuyen el precio de los conjuntos de datos que venden utilizando el método de "pagar para entrenar". Los autores reciben una compensación en función del volumen de sus contribuciones a estos conjuntos de datos.

El valor del "pago por formación" se puede calcular a partir de diversos indicadores de utilidad. Shutterstock, por ejemplo, paga 0,0050 dólares por foto al año, con hasta 0,0038 dólares por foto al año dependiendo de la cantidad de metadatos y otras métricas.

Valor de pago por formación : remuneración de los autores por Shutterstock . Muestra N = 58 (en \$/imagen)



La ventaja de este modelo es su simplicidad de cálculo. Está operativo desde hace dos años y parece tener buena aceptación entre los titulares de derechos.

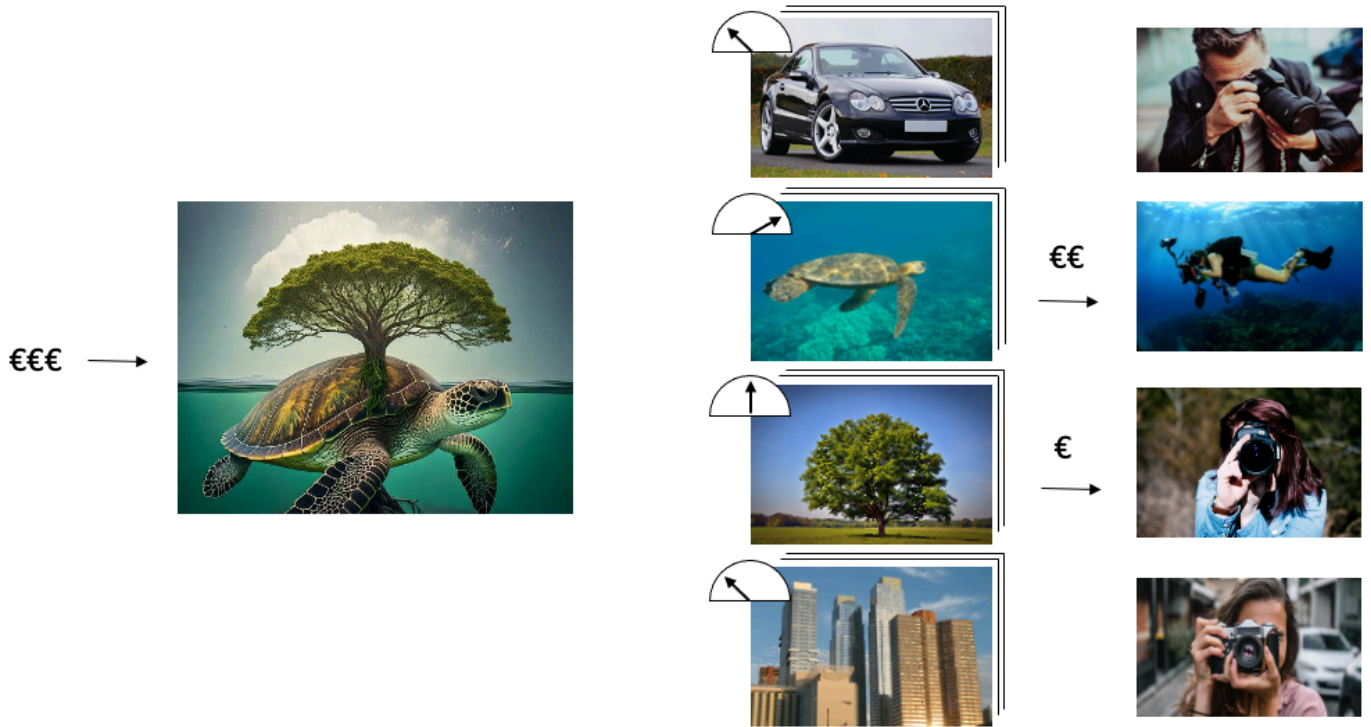
Por otra parte, el nivel de equidad entre autores presenta dos sesgos:

- (13) No se tiene en cuenta la calidad del contenido: un autor que aporta 200 fotografías recibe, en igualdad de condiciones, el doble de remuneración que un colaborador que aporta 100 fotografías.
- (14) No se tiene en cuenta la utilidad del contenido generado: Van Gogh no recibe más que Picasso si los usuarios insertan con más frecuencia "a la manera de Van Gogh" en sus indicaciones.

## Método híbrido Similitud + Pago por capacitación

La startup israelí -estadounidense Bria, que colabora con grandes bancos de imágenes privados (Shutterstock, Getty), está integrada por grandes tecnológicas (Microsoft, Amazon, Nvidia). Como primera aproximación, Bria utiliza, como nuestro demostrador, un método de similitud, con el fin de remunerar a los autores proporcionalmente a la semejanza así medida.

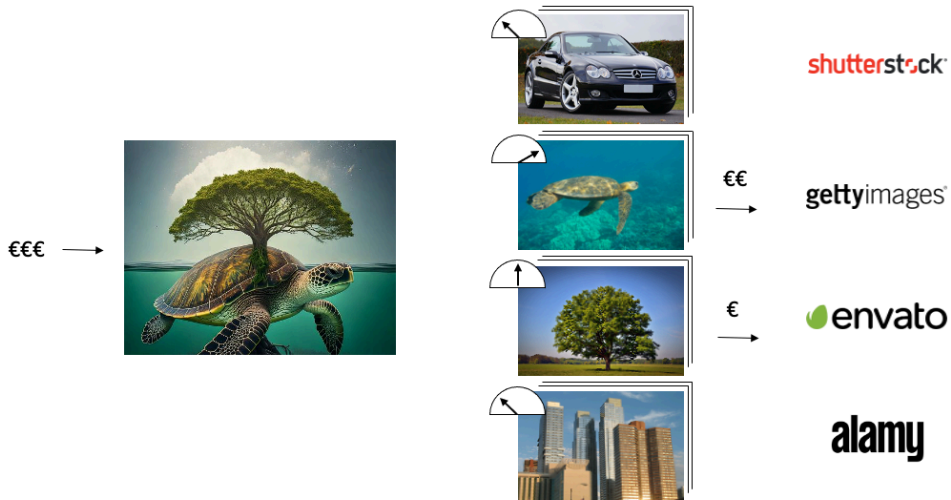
Medir similitudes entre el contenido generado y las fotografías originales para remunerar a los autores



Más precisamente, parece que la suma de las puntuaciones de similitud es calculada por los bancos de imágenes (Shutterstock, Getty, Alamy ...), lo que provoca una remuneración inicial de los bancos proporcional a estas sumas.

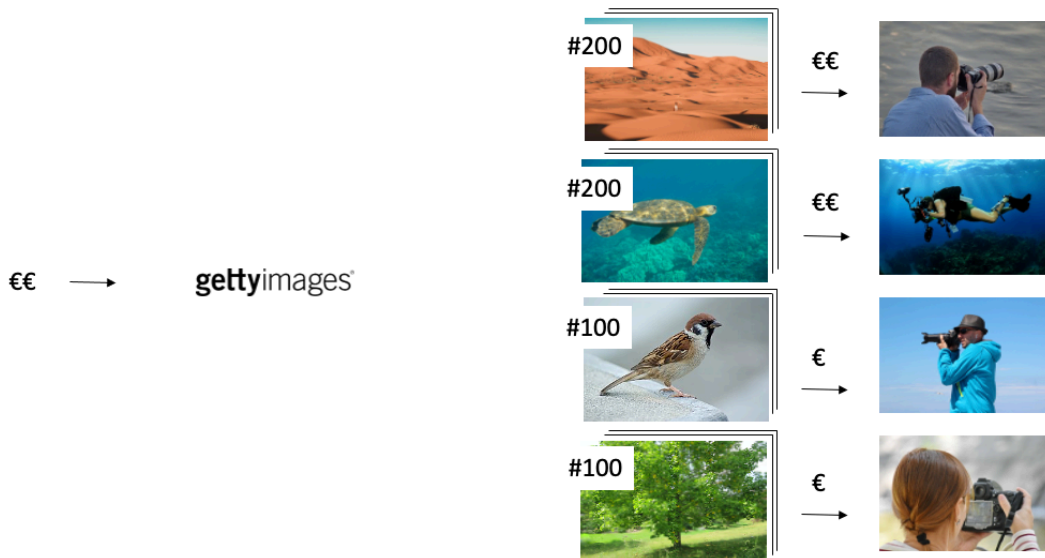


Medir similitudes entre contenidos generados y fotografías originales para remunerar a los bancos de imágenes



Estos bancos distribuyen luego esta primera remuneración entre los autores, en función del número individual de sus obras que contribuyen a los conjuntos de datos vendidos, según el método “ pay to train” .

Número de fotografías a pagar a los autores



Se realizan mejoras adicionales y las imágenes con metadatos de calidad reciben una bonificación.

El método de Bria se realiza por tanto en dos etapas:

- En primer lugar, gracias a un algoritmo de similitud entre las imágenes de entrenamiento y las imágenes generadas
- Luego, gracias a un algoritmo de contribución cuantitativa de los autores a los conjuntos de datos vendidos (“ pay -to-train”)

Esta solución híbrida es:

- económicamente compatible con el modelo de negocio de la IA generativa
- Aceptado por dos años por bancos de imágenes y autores.
- Actualmente se está ampliando a otros tipos de contenidos: 3D, música, voz, imagen animada, vídeo.

Es posible que la startup ProRata , en particular, haya adoptado este método híbrido y lo esté ofreciendo a los medios que ya tienen acuerdos con operadores de IA. Se trata de “tomar el resultado de la IA generativa, ya sea texto, una imagen, música o una película, y descomponerlo en sus componentes, para averiguar de dónde provienen, luego asignar un porcentaje a cada propietario de derechos de autor y pagarles respectivamente .” Las patentes públicas de su empresa matriz no aportan más detalles.

Este método híbrido todavía tiene algunos defectos:

- No está demostrado que esta solución se adapte a todo tipo de contenidos, por ejemplo juegos.
- No hay garantía de que la similitud no sea casualidad. En otras palabras, a un autor se le podría pagar aunque sus obras no fueran útiles para el contenido generado, y viceversa.
- El grado de contribución de una imagen original se mide aproximadamente. En el ejemplo anterior, al fotógrafo del desierto se le paga lo mismo que al fotógrafo de tortugas marinas porque su número de contribuciones al conjunto de datos de Getty Images es el mismo, mientras que las fotos del desierto no tienen ninguna utilidad para el contenido generado.

Cabe señalar que:

- Incluso si la distribución de la compensación por un contenido generado en particular es aproximada, su equidad probablemente mejora con la cantidad de contenido generado.
- No es necesaria la identificación individual de los autores por parte de Bria, sino solo la identificación de los conjuntos de datos de entrenamiento
- El algoritmo de distribución en sí no es público. Es posible que en el futuro los autores exijan más transparencia, pero también es probable que, a la inversa, sea necesaria una dosis de secretismo para evitar sesgos oportunistas.
- Tampoco conocemos la tasa media de remuneración de los autores en relación a la facturación de los contenidos generados que, a nuestro juicio, podrían publicarse.

## Método de similitud de conjuntos de datos

La startup Sureel ofrece a los músicos una herramienta que mide el nivel de influencia que tiene su trabajo, medido en porcentaje, en una pista generada por IA.

El algoritmo, que describe en su patente '949, predice la remuneración de un autor en función de la influencia de todas sus obras, medida por ejemplo por la diferencia entre los pesos del modelo base y los de su ajuste fino por este conjunto. Esta variación pretende diluir el sesgo de similitud a través de un mayor número de obras comparadas.



## Métodos que se están probando actualmente

Más allá de estas pocas variaciones, los investigadores compiten en ingenio para probar nuevos métodos:

Búsqueda de modelos de compensación justos

NEGOCIO	MÉTODO	EQUIPO DE INVESTIGACIÓN	FECHA	MOTIVACIÓN
IA ABIERTA Universidad de Princeton Universidad de Columbia Universidad de Harvard Universidad de Pensilvania Universidad de California Berkeley	En forma de shapley	<b>Una solución económica a los desafíos de derechos de autor de la IA generativa</b> Jiachen T. Wang, Zhun Deng, Hiroaki Chiba-Okabe, Boaz Barak y Weijie J. Su	14 de abril de 2024 9 de septiembre de 2024 (v4)	Los valores de Shapley proporcionan un enfoque basado en principios para distribuir las ganancias en función de la utilidad.  La utilidad ayuda a medir hasta qué punto las fuentes de datos son responsables de generar contenido. El valor es bajo si es poco probable que el modelo contrafactual genere el mismo contenido que el modelo implementado, y viceversa.  La utilidad puede interpretarse como la compensación total que todos los miembros merecen colectivamente por proporcionar sus datos para entrenar el modelo de IA generativa.
Universidad Tecnológica de Virginia Instituto Tecnológico de Massachusetts	Datos de ejecución de Shapley	<b>Datos de Shapley en una sola ejecución de entrenamiento</b> Jiachen T. Wang, Prateek Mittal, Dawn Song y Ruoxi Jia	29 de junio de 2024	Boaz Barak, profesor de seguridad ("superalineación") en Harvard, que también trabaja en OpenAI, no pierde de vista el método ideal de Shapley, incluso si eso significa comprometer la precisión.  Su equipo de investigadores demuestra en la vida real cómo calcular el valor Shapley de conjuntos de datos como Wikipedia o Github para un modelo de IA de propósito general entrenado a partir de un conjunto de datos tipo GPT.  Según este equipo, alinear la remuneración con las contribuciones cuantificables de cada conjunto de datos garantiza la mejor equidad en la distribución de regalías. Ella indica su preferencia por la variante Shapley de datos in-run, menos costosa.
	Modelado de entrenamiento	<b>TRAK: Atribución del comportamiento del modelo a escala</b> Parque Sung Min, Kristian Georgiev, Andrew Ilyas, Guillaume Leclerc, Aleksander Madry, MIT	3 de abril de 2023	Aleksander Madry, profesor del MIT, es responsable del razonamiento de IA en OpenAI. La solución Trak se basa en una representación aproximada del modelo de IA para describir su evolución durante el entrenamiento y relacionar las predicciones con los datos de entrenamiento que las influyeron.  Esta solución ha demostrado ser útil con clasificadores de imágenes entrenados en ImageNet, modelos de visión-lenguaje (CLIP) y modelos lingüísticos (BERT y mT5).

NEGOCIO	MÉTODO	EQUIPO DE INVESTIGACIÓN	FECHA	MOTIVACIÓN
ADOBE Universidad Carnegie Mellon Universidad de California en Berkeley	Similitudes entre conjuntos de datos	<b>Evaluación de la atribución de datos para modelos de texto a imagen</b> Sheng-Yu Wang, Alexei Efros , Jun-Yan Zhu, Richard Zhan	8 de agosto de 2023	Dado que es probable que el método de similitud revele semejanzas coincidentes, este sesgo se puede diluir extendiendo el método a dos conjuntos de imágenes, de entrenamiento y generadas. Los investigadores de Adobe calculan así la contribución del primer conjunto, el conjunto de datos , a un conjunto de contenido sintético.
	Marca de agua	<b>ProMark : marca de agua de difusión proactiva para atribución causal</b> Vishal Asnani , John Colomosse y otros.	14 de marzo de 2024	El tatuaje demuestra la causalidad entre el contenido formativo y el contenido sintético. Esto permite que la atribución legal o los pagos a los creadores se realicen con confianza.
	Desaprender	<b>Atribución de datos para modelos de texto a imagen mediante el desaprendizaje de imágenes sintetizadas</b> Sheng-Yu Wang, Aaron Hertzmann , Alexei A. Efros , Jun-Yan Zhu, Richard Zhang	Junio de 2024	Este método desaprende la imagen sintetizada entrenando las imágenes olvidadas por proxy, identificando aquellas con desviaciones de pérdida significativas después del proceso de desaprendizaje y etiquetándolas como influyentes.
HUAWEI Universidad de Columbia Británica Universidad Simon Fraser Instituto Real de Tecnología KTH	Shapley federado	<b>Mejorar la equidad en la valoración de datos en el aprendizaje federado horizontal</b> Zhenan Fan, Huang Fang, Zirui Zhou, Jian Pei, Michael P. Friedlander, Changxin Liu y Yong Zhang	23 de mayo de 2022	La motivación de los propietarios de datos depende en parte de la equidad de la colaboración y la recompensa en el aprendizaje federado.

## Variantes y desarrollos

En todos los métodos mencionados:

- Los parámetros de remuneración son ajustables en función de arbitrajes sobre indicadores de utilidad: tamaño, fecha, similitud, número de metadatos, éxito comercial independiente, etc.

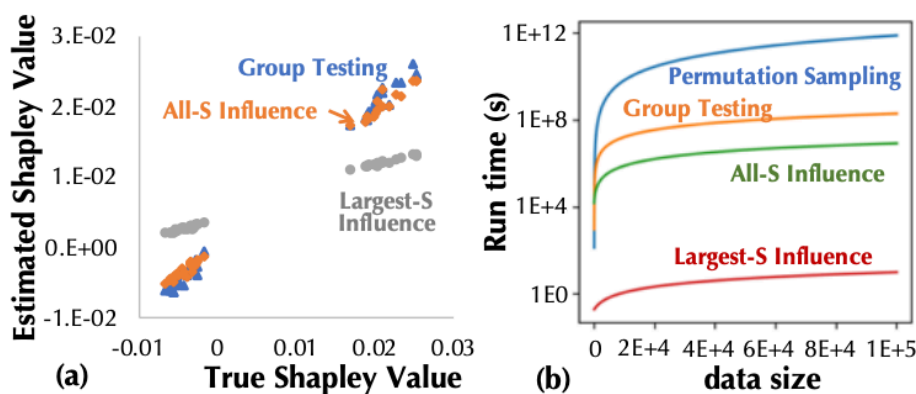
Indicadores de utilidad

INVESTIGADOR, EMPRESA	MEDIOS DE COMUNICACIÓN	INDICADORES DE UTILIDAD
Piao , Chen y otros.	Imagen	Vectores

INVESTIGADOR, EMPRESA	MEDIOS DE COMUNICACIÓN	INDICADORES DE UTILIDAD
Wang, Deng y otros.	Enciclopedia, Publicación académica, Imagen, Código	Modelo contrafactual (S) Contenidos generados por el modelo $x(\text{gen})$ y el modelo contrafactual
Wang, Mittal y otros.	Enciclopedia, Publicación académica, Imagen, Código, Correo electrónico, Foro	Mantener el conjunto de puntos de datos de validación $Z(\text{val})$
Deng, Zhang y otros.	Música	Velocidad media de los acontecimientos Altura media de todos los eventos Suma de las diferencias de tiempo de todos los eventos
Shutterstock + Bria	Imagen, imagen vectorial	Puntuación de similitud Ingresos por conjunto de datos Volumen de datos por conjunto de datos y por autor Volumen de metadatos Papel desempeñado en el desarrollo de los modelos originales Papel desempeñado a través de pagos de regalías vinculados a futuras actividades de licencias generativas
Adobe Stock + Bria	Imagen, imagen vectorial	Puntuación de similitud Admisibilidad de imágenes Fecha de la imagen Número de descargas

- El algoritmo es ajustable entre precisión y velocidad de ejecución. El valor de Shapley, en particular, se puede calcular
  - con un número arbitrario de contenidos generados
  - con granularidad arbitraria (por conjunto de datos , subconjunto de datos , hasta datos individuales)
  - con funciones similares

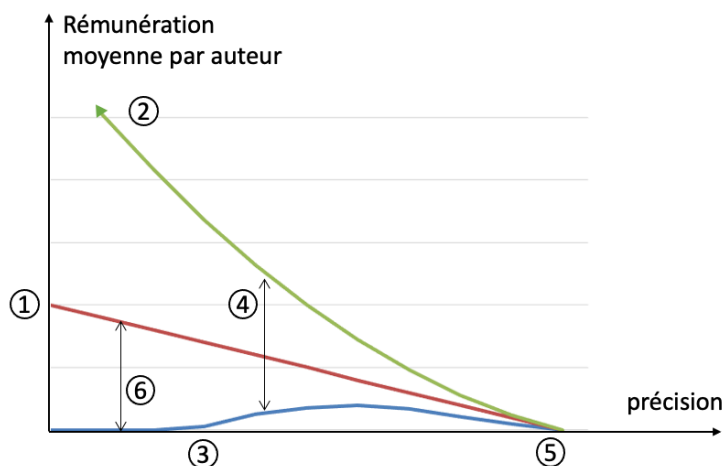
Tres funciones que aproximan el valor de Shapley



En esta etapa del estado del arte, el campo de investigación es suficientemente amplio como para esperar encontrar soluciones de remuneración, por uso, que optimicen la relación entre equidad, costo y precisión.

Impacto del coste de precisión en la remuneración.

- (25) Con precisión cero, la remuneración por autor es igual a la recaudación total dividida por el número de autores.
- (26) Con precisión cero, la remuneración potencial máxima de un autor es igual a la recaudación total.
- (27) Con precisión cero, la remuneración potencial mínima de un autor es igual a cero
- (28) Brecha de incertidumbre
- (29) Cuando el costo de la precisión excede la recolección total, la brecha de compensación e incertidumbre son cero
- (30) Con una recaudación constante, la remuneración media por autor disminuye con precisión



## 7 – Aceptabilidad

**JDN**  
JOURNAL DU NET

**#FairlyShare contre le péché originel d'Internet** – septembre 2014, par Vincent Lorphelin, Gilles Babinet et al.

Ce qu'ils dénoncent ? Les plate-formes qui reposent sur l'industrialisation du travail gratuit, en dehors des règles de l'appel d'offre, du concours ou du bénévolat. Ils emboîtent ainsi le pas à [...] la levée de boucliers qui a suivi le rachat du journal de crowdsourcing Huffington Post par AOL pour 315 millions de dollars. [Lire](#)

método de “ pagar para entrenar” es actualmente el más extendido para remunerar a los autores cuyas obras se utilizan para entrenar modelos de IA de propósito general. A pesar de su parcial transparencia y equidad, aún no ha dado lugar a demandas judiciales ni peticiones, como se vio, por ejemplo, durante la concientización sobre el “trabajo gratuito” o la venta del Huffington Post (ver recuadro).

Los bancos de medios utilizan modelos de “ pago por formación” e inteligencia artificial que utilizan sus conjuntos de datos

FECHA	MEDIOS DE COMUNICACIÓN	PROVEEDOR DE CONJUNTO DE DATOS	PROVEEDOR DE MODELOS DE IA
25/10/22	Imagen, foto	Shutterstock	IA abierta
01/12/23	Imagen, foto	Shutterstock	Meta , Google , Amazon , Apple
12/03/23	Imagen, foto	Shutterstock	LG ( Exaone )
21/03/23	Imagen, foto	Imágenes Getty	Nvidia
29/05/23	Avatar	Arte ascendente	
11/07/23	Imagen, foto	Shutterstock	IA abierta
08/08/23	Imagen, foto	Shutterstock	Nvidia (Picasso)
09/07/23	Imagen, foto	Imágenes Getty Alamy Bria	

FECHA	MEDIOS DE COMUNICACIÓN	PROVEEDOR DE CONJUNTO DE DATOS	PROVEEDOR DE MODELOS DE IA
13/09/23	Música	Audio Sparx	Estabilidad IA
13/09/23	Imagen, foto	Adobe Stock (Firefly)	
26/09/23	Imagen, foto	Imágenes Getty (Getty Gen AI)	
23/04/10	Imagen, foto	Canva (Estudio Mágico)	IA abierta (Dall-e) Google (Imagen)
21/02/24	Imagen, foto	Superstock k Envato	
24/11/04	Imagen, foto	OjoEm Selección libre Fotobucket	
27/06/24	Música, Voz	Música universal* Entretenimiento de Sony Registros Warner	Google ( YouTube )
30/07/24	3D	Shutterstock	Nvidia (Edify)

\*experimentación

Los contratos de compra de conjuntos de datos por parte de los bancos de medios han ido acompañados de discursos que promueven la remuneración “respetuosa, responsable, segura y justa” de los autores.

Discurso sobre la equidad en el pago de la formación

PLATAFORMA	STOCK Y CONJUNTO DE DATOS	DECLARACIONES
Nvidia	Getty, Shutterstock, Adobe	Nvidia : “Hemos creado una plataforma que permite a nuestros socios capacitarse a partir de datos con licencia, por ejemplo de Getty, Shutterstock o Adobe”, dijo Huang. “Respetan a los propietarios del contenido. Los datos de entrenamiento provienen de esta fuente y cualquier beneficio económico que se derive de ellos puede volver a los creadores. »
	Adobe	Adobe: Desarrollamos IA generativa de manera responsable, con los creadores en el centro . Adobe: cree con confianza sabiendo que su resultado es de la más alta calidad, creado con plantillas diseñadas para ser seguras para uso comercial.

PLATAFORMA	STOCK Y CONJUNTO DE DATOS	DECLARACIONES
	Getty	Getty: Generative AI de iStock es una herramienta de IA generativa asequible y comercialmente segura [...] sin el temor de que algo legalmente protegido se haya deslizado en el conjunto de datos y pueda terminar en su trabajo.
Meta	Shutterstock	Shutterstock : Experiencia en la construcción de un ecosistema a gran escala que recompensa y conecta a los contribuyentes con los creadores.
LG Exaone	Shutterstock	<p>Shutterstock : La propiedad del contenido generado por IA no puede atribuirse a un individuo y, en cambio, debe compensar a los numerosos artistas que participaron en la creación de cada nueva pieza de contenido. Nuestros clientes pueden licenciar imágenes de IA de forma segura y legal para sus propias necesidades sin preocupaciones. También compensamos a los colaboradores cuyos trabajos entrenan nuestros modelos, para que usted también pueda utilizar nuestra IA con la conciencia tranquila. Nuestro generador de imágenes de IA sigue siendo la función segura, fácil de usar y diseñada de manera responsable que conoce y ama.</p> <p>LG: El modelo de IA está entrenado con millones de imágenes de alta resolución y metadatos de Shutterstock y convertirá indicaciones de texto en imágenes. Los colaboradores cuyos trabajos se utilizaron para entrenar el modelo serán compensados a través del Fondo de colaboradores de Shutterstock y también serán compensados cada vez que los clientes creen y licencien contenido generativo nuevo que utilice su propiedad intelectual.</p>
IA abierta	Shutterstock	OpenAI : Los datos que obtuvimos bajo licencia de Shutterstock fueron esenciales para entrenar DALL-E. Garantizar que la economía de los creadores siga siendo dinámica es una de las principales prioridades de OpenAI . Los escritores, artistas, compositores y otros creadores han contribuido inconmensurablemente a las sociedades a lo largo de la historia de la civilización [...] OpenAI no quiere reemplazar a los creadores .
	Lienzo	Canva : el convertidor de texto a imagen de Magic Media, DALL·E de OpenAI e Imagen [...] de Google Cloud garantizan que sus modelos de IA sean justos y éticos. “Compartimos el valor que desarrollamos en Canva con los creadores que han estado con nosotros durante todos estos años. Con el nuevo Fondo de Creadores , tienen acceso a un fondo de regalías que coincide con las creaciones de IA. »
Nube de Google	Lienzo Shutterstock	Google Cloud: Shutterstock también se convirtió en un innovador líder al ser el primero en lanzar un generador de imágenes de IA de origen ético, ahora mejorado con Imagen en Vertex AI.
Google Youtube	Unión Soviética	YouTube, subsidiaria de Google, está tratando de llegar a acuerdos con grandes compañías musicales para utilizar inteligencia artificial. Quiere garantizar que " los titulares de derechos musicales reciban una compensación por sus contribuciones a los datos de entrenamiento". UMG y YouTube están en conversaciones para licenciar las voces y melodías de los artistas para entrenar modelos de IA.
Amazon SageMaker	Alamy Getty Envato	<p>Alamy : En el corazón del espíritu del programa Fair Diffusion se encuentra el compromiso de remunerar de manera justa a fotógrafos, artistas y creadores.</p> <p>Getty: Nos adherimos a los más altos estándares éticos y respetamos la propiedad intelectual y los derechos de privacidad de los demás.</p>
Sony		<p>El presidente de Sony dijo: "Los nuevos productos y negocios creados a partir de la música deben desarrollarse con el consentimiento del propietario y con una compensación y crédito adecuados. Es fundamental entender por qué se realiza el entrenamiento de modelos de IA, qué productos se desarrollarán como resultado y cuál es el modelo de negocio que permitirá monetizar el uso de la obra del artista.</p> <p>“Si un modelo de IA generativa se entrena con música con el propósito de crear nuevas obras musicales que compitan en el mercado de la música, entonces el entrenamiento no es un uso justo. En este caso, la formación no puede tener lugar sin el consentimiento, el crédito y la compensación de los artistas y los titulares de derechos. »</p>

## 8 – Patentes

Al preparar este informe, identificamos las siguientes patentes:

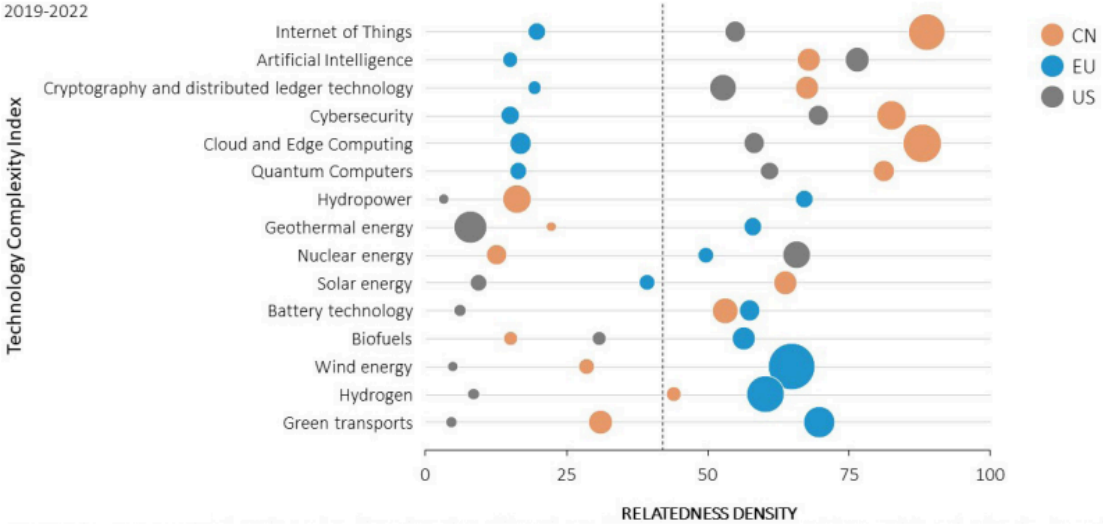
Patentes relativas a la remuneración de trabajos de entrenamiento de IA generativa

TÍTULO	ÁRBITRO	PAÍS DEL DEPOSITANTE	ENTREGADO	PIONERO
Sistemas y métodos de trabajo creativo sistemas colaborativos	US9100444	A NOSOTROS	✓	✓
Valoración del capital de las partes interesadas en proyectos colaborativos	US10776889	A NOSOTROS	✓	✓
Atribución de contenido visual generado a ejemplos de capacitación	US12033372	A NOSOTROS	✓	
Atribución basada en modelos para contenido generado por inteligencia artificial	US12013891	A NOSOTROS	✓	
Generando vectores de autor	US10599770	A NOSOTROS	✓	
Atribución basada en entradas para contenido generado por inteligencia artificial	US2024419949	A NOSOTROS		
Atribución basada en resultados para contenido generado por inteligencia artificial	US2024419720	A NOSOTROS		
Sistema y método de distribución de tarifas	US2023401550	A NOSOTROS		
Generación, uso y atribución de contenidos	WO2024166091	A NOSOTROS		
Facilitar la generación y el uso de contenido visual	WO2023084512	A NOSOTROS		
Método y sistema de selección de clientes en línea basado en el aprendizaje federal, medio y equipo	CN118446287	Porcelana	✓	
Generación de índices de imágenes en función de las similitudes de las características de las imágenes	CN103207879	Porcelana	✓	
Método y sistema para evaluar el grado de contribución de las características locales de la imagen a las características generales	CN111652239	Porcelana	✓	
Método y sistema para cuantificar las contribuciones de los clientes en el aprendizaje federado	WO2023036184	Porcelana		
Método de procesamiento de datos de distribución de dispositivos móviles de aprendizaje federal basado en el valor de Shapley	CN114912626	Porcelana		
Método de excitación del aprendizaje federal basado en el valor de cooperación de Sharpley	CN115965092	Porcelana		
Métodos y sistemas para el aprendizaje federado con predictores locales	WO2024031564	Porcelana		
Método y sistema de colaboración	Estados Unidos 9723059	UE	✓	✓
Método para un autocodificador explicable y una red generativa adversarial explicable	US11948083	UE	✓	
Arquitectura para el aprendizaje de refuerzo explicable	US11455576	UE	✓	
Trazabilidad de contribuciones en múltiples etapas en entornos de creación colectiva	US2015363743	UE		
Método y dispositivo para premiar la creación colaborativa	EP2958068	UE		



Esta lista muestra que, en lo que respecta a la remuneración de los titulares de derechos, los europeos no se quedan atrás de China y Estados Unidos, lo que contrasta con su posición de outsider en lo que respecta a la inteligencia artificial en general y a la IA generativa en particular.

**The EU's position in complex (digital and green) technologies**



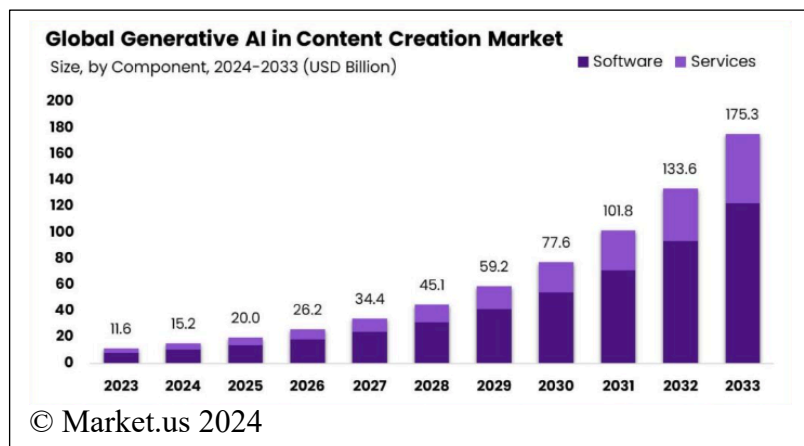
**Table 7 Top patent owners in GenAI modes (companies), 2014-2023**

	Images/ video	Text	Speech/ voice/ music	3D image models	Molecules/ genes/ proteins	Software/code	Other modes
Tencent Holdings (China)	607	565	551	102	57	41	464
Ping An Insurance (China)	262	600	599	26	33	24	223
Baidu (China)	395	465	441	81	26	18	166
IBM (US)	101	274	168	13	37	42	132
Alibaba Group (China)	142	213	144	36	2	10	133
Samsung Electr. (Republic of Korea)	173	140	226	34	10	5	57
Alphabet/Google (US)	138	107	200	25	15	15	72
ByteDance (China)	173	82	112	21	6	7	90
Microsoft (US)	78	194	151	18	6	22	42
BBK Electronics (China)	125	58	117	12	0	4	116
Netease (China)	78	76	80	37	0	11	102
NTT (Japan)	37	100	169	11	0	3	62
Huawei (China)	104	84	72	24	1	4	102
China Mobile (China)	79	67	83	10	0	7	89
State Grid (China)	75	54	31	6	1	2	144
Adobe (US)	190	79	34	30	0	9	20
Sony Group (Japan)	87	35	105	42	0	1	42
Siemens (Germany)	124	25	8	10	3	4	59
Ant Group (China)	45	82	29	3	0	6	59
Industrial and Commercial Bank of China (China)	40	62	49	5	0	16	48

Source: WIPO, based on patent data from EconSight/IFI Claims, April 2024.

**Perspectivas**

## 9 – Mercado

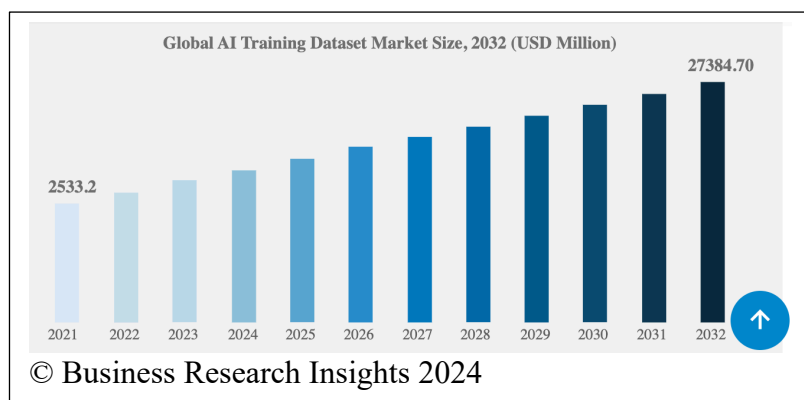


Solo para el segmento de creatividad, se estima que el tamaño del mercado global de IA generativa alcanzará los USD 11.600 millones en 2023 y superará los USD 100.000 millones en 2031.

Para dar una idea de la escala, una tasa de remuneración justa del 15% representaría más que la recaudación actual de las OGC (Organizaciones de Gestión Colectiva) por derechos artísticos (música, audiovisual, espectáculos en vivo, literatura y artes visuales).

Las empresas especializadas en conjuntos de datos de entrenamiento de IA organizan la recopilación, anotación, curación de datos y su adaptación a las necesidades de los algoritmos de aprendizaje.

El mercado global de conjuntos de datos de entrenamiento de IA fue de 2.500 millones de dólares en 2021 y se espera que alcance los 27.000 millones de dólares en 2032.



### Modelos de compensación para autores de datos de IA generativa

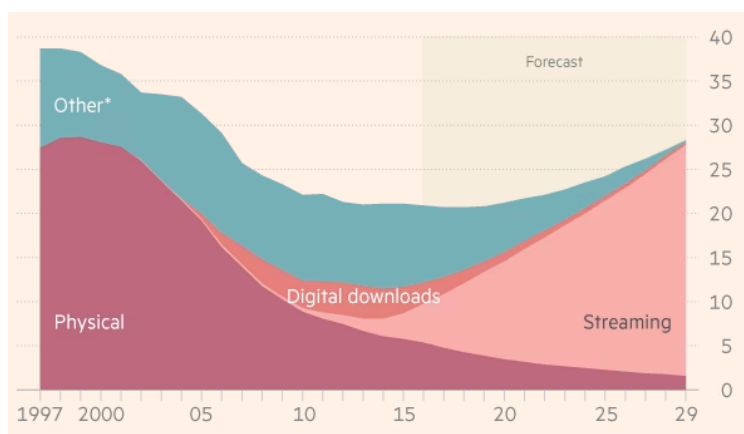
MODELO DE NEGOCIO	EMPRESAS	COMPENSACIÓN DEL AUTOR
Raspado (1)	OpenAI , Midjourney , Estabilidad, MistralAI	No
Uber de datos (2)	Mturk , ClickWorker , Gigwalk , Premisa	Pago en el momento del pedido
Bancos de contenidos (3 )	Shutterstock, Getty, Canva, Adobe	Pagar para entrenar
Grupos de medios (4)	Axel Springer, Financial Times, Stack Overflow, Reddit, Le Monde	No
Empresas de autores (5)	el gremio de autores , Creado por humanos, IA nativa humana, entrenada de manera justa, Alianza de proveedores de conjuntos de datos	Salario justo

- Scraping : DALL·E, por ejemplo, se entrenó primero con cientos de millones de imágenes subtituladas de Internet. Esta práctica, que es mayoritaria, es motivo de numerosos litigios.
- Uber de datos: los creadores de modelos de IA han comenzado a cubrir riesgos y proteger las cadenas de suministro de datos, tanto a través de una próspera industria de corredores de datos como mediante acuerdos con propietarios de contenido que han surgido para satisfacer la demanda. Está surgiendo una industria de empresas especializadas en datos de inteligencia artificial que crean redes de trabajadores contratados a corto plazo para producir contenido personalizado a pedido. Los precios varían entre \$1 y \$2 por fotograma, entre \$2 y \$4 por video corto y entre \$100 y \$300 por hora para películas más largas. La tarifa de mercado para el texto es de \$0,001 por palabra. Los propietarios de las fotografías, podcasts y datos médicos reciben entre un 20 y un 30% del importe total de la transacción.
- Bancos de contenidos: Shutterstock , Getty, Canva o Adobe cuentan con grandes bases de medios suministradas por los autores, a quienes se les paga según los conjuntos de datos vendidos. Shutterstock , el líder en imágenes, ha vendido licencias a proveedores de IA por más de 100 millones de dólares, o el 12% de sus ingresos totales, y se espera que crezca un 40% en 2024. Estas licencias de imágenes, vídeos y música formarán un mercado direccionable de 10 mil millones de dólares en 2030, creciendo anualmente un 22% durante ese período.
- Grupos de medios: Los propietarios de bases de datos otorgan licencias a los conjuntos de datos , sin compensación adicional <sup>50</sup>los autores
- Sociedades de autores: los autores quieren poder definir sus opciones de licencia para la IA generativa. Las ideas están floreciendo como la del *Gremio de Autores* con la plataforma *Created by Human* , como *Human Native AI* , *Fairly Capacitado* , *Alianza de proveedores de conjuntos de datos* o la *Comisión de IA* en Francia. En esta etapa no se han definido los modelos económicos ni se han cuantificado las perspectivas de desarrollo.

## Propuestas de licencias para obras de IA generativa

INICIATIVA	PRINCIPALES PROPUESTAS
El gremio de autores	<ul style="list-style-type: none"> <li>- Establecimiento de acuerdos para la utilización de obras</li> <li>- Pago de regalías por el uso de obras protegidas</li> <li>- Transparencia en los datos utilizados</li> </ul>
Creado por Human	<ul style="list-style-type: none"> <li>- Etiqueta “ Creado por humanos ” para obras</li> <li>- Plataforma que permite a los autores autorizar y controlar licencias</li> </ul>
IA nativa humana	<ul style="list-style-type: none"> <li>- Mercado para conjuntos de datos de entrenamiento de IA generativa</li> <li>- Seguimiento de permisos para cada contenido utilizado</li> </ul>
Bastante entrenado	<ul style="list-style-type: none"> <li>- Etiqueta “ Bastante Entrenado » para IA generativa</li> <li>- Contratos estandarizados para el uso de obras en IA</li> </ul>
Alianza de proveedores de conjuntos de datos	<ul style="list-style-type: none"> <li>- Remuneración justa</li> <li>- Contratos estandarizados ( Pay -to-train, regalías sobre facturación, suscripción, licencia por sector)</li> <li>- Transparencia en el uso de los datos de formación</li> <li>- Autorizaciones individualizadas por autor</li> </ul>
Comisión de IA	<ul style="list-style-type: none"> <li>- Plataforma pública</li> <li>- Intermediación de la Biblioteca Nacional de Francia y el Instituto Nacional del Audiovisual</li> </ul>

Haciendo un paralelo con el mercado de la música, actualmente estamos en la fase “ Napster -IA”, a la espera de soluciones equivalentes a Spotify y Apple Music. Sus modelos de negocio limpiaron el mercado a partir de 2012, donde antes el streaming y la piratería se confundían.



© Financial Times, Goldman Sachs

De igual modo, creemos que el mercado potencial de valorización de obras mediante IA generativa es mayor que el de las OGC, y representa una oportunidad para los titulares de derechos. De forma simétrica, los operadores de IA generativa podrían acceder a bases de contenido de calidad.

Este potencial, que por el momento beneficia a segmentos alternativos del mercado, se liberará con la propuesta de nuevos modelos económicos del tipo de

las "sociedades de autores" y su exploración voluntaria por parte de los interesados.

A pesar de su carácter embrionario, la multiplicación de estas iniciativas abre una nueva perspectiva de “propiedad líquida”.

## 10 – Hacia la propiedad líquida

### Interview de Vincent Lorphelin

Le Soir (Bruxelles), Janvier 2024



"Sam Altman y Elon Musk dice que la IA generativa hará que el trabajo quede obsoleto, así que prepárense para un ingreso universal. Esto concuerda con la tesis del fin del trabajo transmitida por Jeremy Rifkin hace unos años. Esto también se relaciona con lo de la singularidad, que sugiere que dentro de unos años, la IA general ... ir Superar a los humanos en todos los ámbitos. La prueba es que siempre hemos dicho que la IA nunca entrará en las profesiones creativas (nunca será autora, diseñadora...). Excepto que desde el año pasado, estamos allí. Las puertas del bastión están derribadas. [...]"

Por el contrario, el ADN europeo es la economía social de mercado, es la economía que crea la sociedad. El trabajo tiene sentido. Históricamente y desde un punto de vista estratégico, la gestión colectiva de derechos está dos veces más desarrollada en Europa

que en Estados Unidos. Por lo tanto, si volvemos a esta cuestión fundamental de interés. En general, para contrarrestar la tesis de la singularidad, el interés general Europa debe promover el modelo descentralizado.

Imaginemos que este modelo de plataforma descentralizada se vuelve sistémico, se convierte en un sistema, de la misma manera que lo hacen hoy las plataformas centralizadas como Gafam. En este caso, los asuntos manipulados, son derechos intelectuales, no mercancías. Este es un valor económico no comercial. El economista Karl Polanyi había argumentado que la economía de mercado había mercantilizado el trabajo, mientras que el trabajo no puede equipararse a una mercancía.

Las plataformas centralizadas intentarán en la medida de lo posible dar un valor de mercado a los derechos intelectuales. En una lógica descentralizada, los derechos se expresarán en porcentajes, no en moneda. Los valores económicos manipulados ya no se expresan en euros ni en dólares. Es una economía que no se basa en el dinero sino en la propiedad. Para trazar un paralelo con lo que Émile Zola llamó "dinero líquido" (con la transición de una economía donde la riqueza ya no provenía de la propiedad de la tierra sino de la circulación del dinero), se propone la noción de "propiedad líquida", donde el capital de trabajo de la economía es la circulación de los derechos de propiedad intelectual.

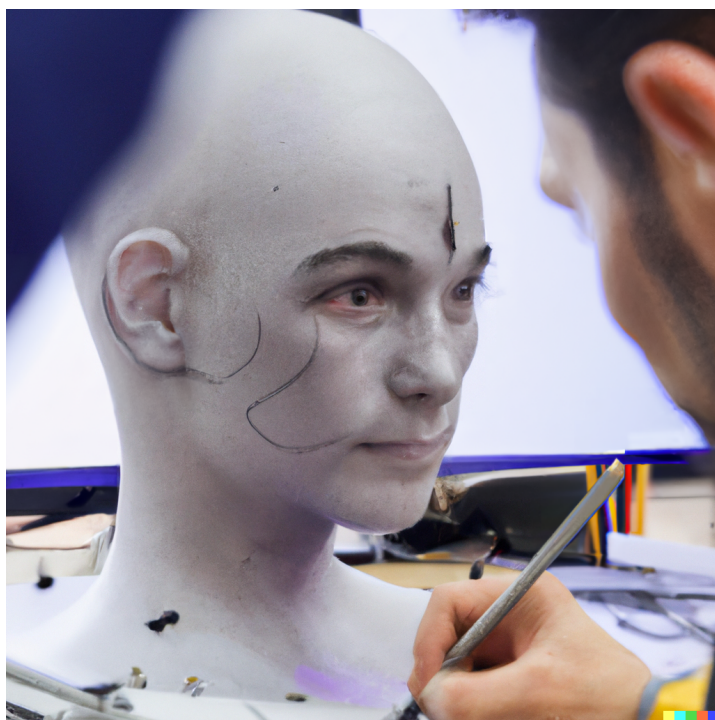
Hoy en día hemos financiarizado todo tanto que una empresa presenta sus cuentas como si estuviera en venta cada día, porque todo lo mide en dinero. Sin embargo, existen activos intangibles, fuera de balance, que expresan valores que no se expresan en dinero. La riqueza de la empresa va más allá de lo que expresa el dinero. Cuando manipulamos las propiedades de los líquidos, nos vemos obligados a encontrar otros indicadores. "Poco a poco vamos avanzando hacia una sociedad que hace solvente una riqueza de otro orden". (leer más [aquí](#)).

Así como el modelo de plataforma de Uber inspiró el concepto de uberización, llamamos “sacémización” al fenómeno de la difusión del modelo económico de las sociedades de derecho de autor, que valorizará el capital humano a través de la propiedad líquida.

## 11 – Promover la inteligencia humana

Brad Smith, presidente de Microsoft, está cometiendo un error moral contra la empresa. Hablando sobre la IA generativa como ChatGPT, dice: “Todos tenemos derecho, según la ley de derechos de autor, a leer y aprender. Ahora nos preguntamos si podemos permitir que las máquinas aprendan de la misma manera. Creo que hay un imperativo sociedad para hacer esto posible. »

Esta intervención puede parecer trivial. Después de todo, si las máquinas están absorbiendo Internet para crear artículos e ilustraciones a un precio ridículamente bajo, ¿por qué privarnos? ¿No están el New York Times y otros titulares de derechos que se oponen a este uso de su contenido dándole un mal juicio a Microsoft? Si las máquinas no plagian, ¿qué daño hay? Éste es el inaceptable resultado final de este discurso.



### Una vaguedad jurídica ventajosa

las plataformas Gafam recogen una gran cantidad de pequeñas contribuciones y promueven el trabajo gratuito. Vuelven a cambiar de dimensión con la IA, cuya voracidad exige la conquista de nuevos territorios, los de los autores.

Como excepción, el legislador había definido el “uso justo” para autorizar a Google a publicar extractos de contenido privado. Con el resultado del tribunal aún incierto, el ex abogado Brad Smith sabe que la opinión pública jugará un papel importante. Por eso destila la imagen mental adulterada del “derecho de las máquinas a

aprender”. Su “imperativo Social” es sólo una vulgar contestación de la legitimidad de los autores para defender sus territorios.

Él distrae debatiendo cuestiones éticas falsas. ¿Está la inteligencia humana condenada a ser... ¿degradado? ¿Podemos dejar que la IA tome decisiones? Bien podría adoptar la postura de un hombre sabio frente a un Elon Musk, menos educado, ambos son augures de la misma ilusión: la de superar la inteligencia humana.



## **IA: ¿una ilusión de superioridad ?**

La mejor novela nunca reemplazará la experiencia de vida . La mejor “ videollamada ” nunca sustituirá a una reunión real. Por muy sofisticado que sea el modelo , no es el real . Cada uno de nosotros tiene una inteligencia del mundo que siempre superará a la de las máquinas, que no tienen otra inteligencia que la de sus modelos .

Las IA iluminan nuestra inteligencia. Permiten a la policía medioambiental detectar a los contaminadores, al profesor comprender las dificultades de mil alumnos o al laboratorio anticipar los efectos de un medicamento .

Difundir la ilusión de la inminente superioridad de las máquinas es irresponsable. Se trata de negar la obra estrictamente humana, que se inspira precisamente en el mundo y sus modelos , cuando el artista revela algo no dicho, cuando el conductor compensa palancas que responden mal , cuando el médico duda de un diagnóstico demasiado obvio , o cuando el El juez tiene una convicción íntima.

### **La defensa de los autores**

Alimentar la ilusión de una competencia entre inteligencias y atribuir derechos humanos a las máquinas es una falta contra la sociedad . Por el contrario, debemos afirmar la preeminencia de la inteligencia humana y garantizar la integridad de sus territorios.

La ilusión de Brad Smith no tendría sentido si no fuera por la figura más importante del mundo en inteligencia artificial. Conquistada por esta ilusión, la Comisión Interministerial sobre IA afirma que las máquinas superarán a los humanos. Ella defiende , afortunadamente, la justa remuneración de la creación humana , pero sin darse cuenta de que es la apreciación Incluso de esta exactitud que es manipulada .

Europa ya ha decidido con la AI ACT proteger los derechos de autor. Ahora debe establecer el registro de las tierras explotadas por los IA y defender su renta. ( [Leer](#) más en Les Echos del 6 de mayo de 2024)



## 12 – Una solución sacemizadora para Europa

### Entretien avec Jean-Paul Betbèze

Economiste, Membre du Comité scientifique de la Fondation Robert Schuman



### ¿Cuáles deben ser los puntos claves de la “ sacemización ”?

solución “ sacemizadora ” para Europa debería:

- Basarse en la propiedad intelectual de los autores, como las OGC. Es necesario mantener y garantizar la propiedad intelectual de los autores, que de otra manera podría disolverse o desaparecer en la actual revolución de la comunicación. La comunicación se está globalizando, es la base del gran cambio que estamos viviendo, con la multiplicación de mensajes en todas direcciones. Hay que reconocer la contribución de cada uno, a menos que se seque el movimiento global actual con una concentración excesiva en beneficio de ciertos autores, sin que los demás encuentren ninguna ventaja en su trabajo y sus ideas, ningún reconocimiento, ninguna remuneración. Debido a las brechas en la disponibilidad de capital, en gran medida en

beneficio de Estados Unidos , el peligro es una monopolización de las fuentes aparentes de creación.

- Recaudar regalías por la explotación de obras de IA. La “ sacemización ” no sólo permite rastrear, sino también identificar y estimar fuentes, incluso cuando parecen mínimas. Este proceso permite preservar todos los orígenes y, por tanto, todo reconocimiento de las creaciones y mejoras: nada se pierde, todo se conserva.
- Reconocer la creatividad y distinguirla del contenido amateur. En la proliferación de obras, es importante poder destacar aquellas que son originales, resaltarlas y promoverlas.
- Implementar mecanismos de incentivos individuales, de modo que el precio de la creatividad sea competitivo en el mercado y compatible con el costo de producción. El riesgo de multiplicar el trabajo es su banalización, sobre todo porque su coste físico de producción es evidentemente bajo. La facturación vinculada únicamente al coste de producción corre el riesgo de resultar desincentivadora . La gran ventaja de la “ sacemización ” es pasar de una lógica del precio ligada al coste de producción, para pasar a otra lógica, la de los mercados, que se basa en la importancia del éxito, la repetición del mensaje.

### ¿Cuál es la ventaja de un mercado impulsado por la “ sacemización ”?

En esta lógica de mercado, es la demanda medida, de la que surge la demanda anticipada, la que crea los nuevos precios, y por tanto los márgenes. En este contexto, los márgenes sólo pueden aumentar, por lo tanto

los beneficios marginales, por lo tanto las valoraciones. Esta diferencia entre costes bajos, detectados por los motores de búsqueda y valoraciones que tendrán así una base objetiva, actualizada por herramientas de tipo Sacem, es la base de la lógica propuesta. Permitirá destacar las contribuciones de todos, especialmente en un contexto europeo.

### ¿Y para las actividades creativas?

No olvidemos que la competencia monopolística está siempre presente: empuja hacia una reducción del número de competidores, hasta que sólo surge uno: el monopolio. Allí, más que en ningún otro lugar, el ganador se lleva todas las apuestas. La “sacemización” debe oponerse a esto permitiendo una mejor distribución de la remuneración antes de que aparezca un solo monopolista. Esto es evidentemente favorable a la investigación, a la innovación y a la creación, ya que sabemos que la situación de monopolio se vuelve contra el propio monopolista, en una situación de mercado que se seca a largo plazo, porque su dinámica conduce a una caída de sus beneficios. La “sacrificación” respeta así las creaciones y frena la selección despiadada y, en última instancia, destructiva del propio monopolio.

En el mercado de las ideas, la velocidad es mortal, el recuerdo de las aportaciones es decisivo, unido a su justa remuneración. Hoy es posible y, comprensiblemente, más necesario que nunca. Se trata de salvar la creación a través del progreso que ofrece la tecnología. »

Michelle Bergadaà, Professeure émérite à l'université de Genève.

Co-autrice de *Réinventer l'intégrité académique à l'ère de l'intelligence artificielle* (à paraître, EMS, février 2025)



### Usted dice que la Ley de IA no sólo afecta a los autores, sino también a los investigadores, ¿por qué?

MB: “Deberíamos considerar la posibilidad de aplicar la Ley de IA para compensar a los autores de “investigaciones auténticas”” Francia paga mal a sus investigadores, lo que frenaría su fuga al extranjero o al sector privado. A la sociedad le cuesta mucho más aplastar a sus jóvenes investigadores bajo la burocracia que pagarles bien. Muchos se desmotivan y hacen horas extras de docencia en lugar de investigar. Así que si el Estado no puede pagarles adecuadamente, el sistema de producción de artículos, libros e informes debe hacerlo. Ésta debe ser la piedra angular de la estrategia europea de IA generativa. Con la ventaja adicional de un sistema automatizado de detección de fraude.

### ¿No ha existido siempre el fraude?

La crisis de los opioides en EE. UU. [El fentanilo y otros medicamentos relacionados, recetados con receta, por sí solos fueron responsables de aproximadamente 71 000 muertes por sobredosis de drogas en 2021] se produjo debido a un solo artículo falso, mal citado y reimpresso. Esto fue mucho antes de la IA. Con la IA, la difusión de artículos

fraudulentos se realiza principalmente a través de artículos nuevos. Éste es un nuevo modelo de ataque a lo falso.

Hoy coexisten tres temporalidades. En primer lugar, el tiempo instantáneo y emotivo (de Trump y la Generación Z) permite la producción no sólo de información falsa, sino también de autores falsos, igualmente creíbles. En segundo lugar, el presente eterno de las comunidades de antaño mantiene un mundo alternativo donde la tierra es plana y la lejía elimina el Covid . El tercer tiempo lineal está impulsado por la tecnología materialista de Elon Almizcle . La IA es un acelerador del fraude que prospera en los intersticios de estas temporalidades.

### **¿Cómo actuar?**

Las bases de datos Pubpeer y Retraction Watch buscan atrapar a los infractores individuales, uno por uno. Pero pasaron tres años hasta que se completó la investigación y se obtuvo una retractación en los periódicos.

Este fenómeno debe abordarse sistemáticamente con radares de integridad, una orden judicial para que los defraudadores devuelvan el dinero recibido y la exclusión. Bases de datos de IA. El futuro de nuestra sociedad son los jóvenes que educamos. Ayudémosles adoptando una postura firme y remunerando sus creaciones originales a través de la IA generativa. Esto motivaría a muchos investigadores jóvenes y honestos a crear conocimiento. »

# Recomendaciones

## 13 – La solución Controvrse

Sam Altman, fundador de Open AI, quiere tener "nuevos modelos de negocio que permitan a los creadores generar ingresos". »

Hemos visto que el Uber del modelo de datos surge para evitar la degeneración de la IA generativa, pero que amenaza el capital humano. El modelo económico debe apoyarse en las OGC para abrir una perspectiva de revalorización de este capital a través de la propiedad líquida y la sacemización .

El modelo de negocio que recomendamos a Sam Altman y otros operadores de IA generativa es el siguiente:

### Solución a la controversia

RECOMENDACIÓN	IA ABIERTA	OGC
1 - Reserva	Reservar el 15% de la facturación de la IA generativa para una remuneración justa de los autores y titulares de derechos	Proporcionar a los autores una plataforma para oponerse, cuando corresponda, al uso de sus obras por parte de la IA generativa, o para elegir una licencia, obra por obra.
2 – Segmentación	Identificar los principales OGC por segmento de mercado (medio/ conjunto de datos * país)	Reclamar un segmento de mercado (medio/ conjunto de datos * país)
3 - Método	Publicar el método de atribución por OGC, detallar la desambiguación en caso de recuperación	Negociar acuerdos de coordinación o reciprocidad entre OGC
4 – No atribución	Identificar un fondo por país para montos no asignados	Gestión de reclamaciones del fondo
5 - Informes	Publicar el resumen con suficiente detalle  Enviar a cada OGC identificada un informe mensual que contenga los identificadores de los autores, el detalle del uso de sus obras, por contenido sintético, y el valor atribuido.	Envía al operador la lista de obras que no pueden ser utilizadas por la IA generativa
6 - Pago	Distribuir la reserva entre las OGC y los fondos	Distribuir las cantidades recaudadas entre los autores y titulares de derechos

## 14 – Iniciar la negociación

### Pasos

Para llegar a acuerdos entre los operadores de IA generativa y los CMO, recomendamos que las partes interesadas documenten los siguientes pasos:

#### Pasos de la solución de Controv3rse

RECOMENDACIÓN	PROCESO	MEJORES PRÁCTICAS, COMPARABLES
1 - Prototipo	Discutir a partir de una solución técnica a perfeccionar para facilitar la proyección de los interlocutores en un resultado y reducir la complejidad del debate (ver nuestro demostrador)	Demostrador, métodos de prueba y error
2 - Visión	Desarrollar la previsión: descentralización vs. plataformas centralizadas, sacemización vs. Uberizing AI, trabajo de autores vs. Mercantilización, absurdo del “derecho” de la máquina, propiedad líquida, métricas de utilidad	Marco europeo: economía social de mercado, economía de la creatividad, patrimonio y diversidad cultural
3 - Remuneración justa	<p>Especificar la tasa justa de remuneración en función de la medición de la utilidad de las obras de todos los autores para todo el contenido generado. Ofrecemos un valor inicial del 15%.</p> <p>Proponer indicadores de la utilidad de las contribuciones para la atribución autor por autor, y el método para refinar esta atribución en previsión del progreso técnico futuro.</p>	<p>Difundir Directiva 2006/115/CE (13) Derechos conexos (73)</p> <p>Estadísticas de Yacast para Spré</p>
4 – Código de buenas prácticas de la Oficina de Inteligencia Artificial	<p>Observatorio de última generación</p> <ul style="list-style-type: none"> <li>- Métodos: valor de Shapley, pago por entrenamiento, similitud</li> <li>- Arbitrajes: parámetros, granularidad, indicadores de utilidad</li> <li>- Plataformas de opciones y licencias</li> <li>- Identificación de los titulares de derechos y autores</li> <li>- Procesamiento de importes no asignados</li> </ul> <p>Resumen suficientemente detallado:</p> <ul style="list-style-type: none"> <li>- Contenido</li> <li>- Granularidad</li> <li>- Formato</li> </ul> <p>Informes a los titulares de derechos:</p> <ul style="list-style-type: none"> <li>- Reclamaciones de conjuntos de datos por parte de los titulares de derechos</li> <li>- Modalidades de asignación y distribución por el titular de los derechos</li> <li>- Tramitación de recuperaciones</li> <li>- Datos</li> </ul> <p>Contenido secreto: Información no divulgada en resúmenes e informes</p> <p>Normalización progresiva</p>	<p>Código de buenas prácticas contra la desinformación</p> <p>Recomendaciones del gremio de autores</p> <p>Misión de la CSPLA</p> <p>CSI, ISAN</p>
5 – Patentes	Garantizar posiciones francesas y europeas mediante una cartera de patentes dedicadas, en cooperación público-privada	Programa PI de la Alianza INPI

## Salario justo

La compensación justa a través de la IA generativa podría inspirarse en la del Spré . En Francia, esta organización recauda un canon calculado sobre el volumen de negocio de los clubes nocturnos en beneficio de los músicos. Una muestra representativa de 120 de ellos está equipada con una caja “ Yacast ”, financiada por Sacem y Spré . Yacast recopila listas de reproducción de música transmitida y publica estadísticas por artista, sello, género y número de semanas de transmisión. Tras la deducción legal del 9%, el Spré distribuye la tasa entre las OGC (SCPA 50%, ADAMI 25% y SPEDIDAM 25%) y les comunica las estadísticas de Yacast para remunerar a los beneficiarios finales (artistas intérpretes, productores, acciones de interés general).

## Tarifa justa

En su informe de 2020, *El autor y el acto de creación*, Bruno Racine lamentaba que «la remuneración proporcional de los autores sólo excepcionalmente alcanza o supera el 10% de los ingresos de explotación». Atribuyó esta constatación a "la relación que vincula al artista-autor con los actores posteriores (editores, distribuidores, productores) [que] parece profundamente desequilibrada, lo que lleva, en lo que a él respecta, a cuestionar en muchos casos la idea misma de “libertad contractual.” Para remediarlo, quería "reforzar a los artistas-autores colectivamente, mediante la rápida organización de elecciones profesionales que permitan dar contenido y legitimidad al Consejo Nacional de Artistas-Autores que se creará para servir de marco a negociación colectiva con los radiodifusores" para establecer "un sistema contractual de remuneración del tiempo de trabajo del autor, preferente por el patrocinador".

Dado que la idea de revalorizar el valor del trabajo no es aplicable al caso de la IA generativa, proponemos definir la tasa justa como el valor de la utilidad de todas las contribuciones de los autores reportadas a la cifra de negocio del contenido generado. Este es el principio en el que se basa el valor Shapley y el economista Ernst Fehr estima que el valor que aporta la prensa a Google Search es del 14% . Desde la famosa pregunta “¿Te resultó útil este comentario ? ” » lanzado por Amazon, todos los sitios web comerciales mejoran la calidad de su oferta multiplicando las métricas de utilidad, que hoy son de uso común.

Fijar la base imponible en la facturación de los contenidos generados permitiría que la remuneración justa del autor no dependiera de la mayor o menor segmentación vertical del sector, por ejemplo: autor/plataforma/ conjunto de datos /modelo general/ ajuste fino /interfaz/ generación.

## 15 – Licencia – Agradecimientos

Esta nota y sus tablas son accesibles bajo licencia CC-BY, sujeta a la distribución justa de los frutos de su explotación por IA generativa, otorgada por la asociación Controv3rse, RNA W751267044.

Este documento fue traducido automáticamente por Google Translate. En caso de ambigüedad, prevalecerá la versión francesa.

Se inspira en las reflexiones y debates entre los miembros de las misiones ministeriales para la Ley IA , las de los think tanks independientes Controv3rse y el Institut de l' Economie , así como otras personalidades reconocidas por su experiencia. Prolonga el estudio realizado previamente para la misión parlamentaria



Pradal -Rambaud y las columnas publicadas en *Le Monde* y *Les Echos* . Vincent Lorphelin, fundador y copresidente de estos think tanks, autor de esta nota, les agradece calurosamente por sus ideas y la calidad de estos debates.

#### Presidentes y Relatores de las Misiones Parlamentarias y Ministeriales:

Alexandra Bensamoun , profesora de Derecho en la Universidad París Saclay, ex miembro de la Comisión de Inteligencia Artificial

Joëlle Farchy , profesora universitaria, directora de m2ecn y de la Cátedra PcEn Universidad París 1 Panthéon-Sorbonne, miembro de la CSPLA

Philippe Pradal , diputado por la tercera circunscripción de los Alpes Marítimos, Asamblea Nacional

Stéphane Rambaud, diputado por la tercera circunscripción del Var, Asamblea Nacional

Lionel Ferreira, Maestro de Solicitudes del Consejo de Estado

Julie Groffe -Charrier, profesora de Derecho privado en la Facultad Jean-Monnet

Bastien Blain, profesor adjunto de economía, científico de datos y del comportamiento, Universidad París 1 Panthéon-Sorbonne

#### Expertos y emprendedores:

William Bailey  
Cofundador y presidente de Bolero Music

Director de Marketing, Comunicación e Innovación, Keyrus

Emmanuel Benazera  
Presidente y director ejecutivo de Jolibrain

Alicia Coatalem  
Cofundador de CogNeed y profesor asociado de la Universidad Paris Dauphine PSL

Jean-Paul Betbèze  
Economista , Betbèze Conseil

Cristophe Collet  
Fundador, AskLocala

Michelle Bergadaá  
Presidente del Instituto de Investigación y Acción sobre el Fraude Académico y el Plagio

Federico Dayan  
Fundador, Exactcure

Michel Bokobza  
Director de misión, Collège de Paris

Guillermo Desveaux  
Cofundador de Aleia y administrador de AI Cargo Foundation

Laurent Cervoni  
Director de Investigación e Innovación, Talan

Jonathan Dory  
Fundador , Live Crew

Charles-Éric de la Chapelle  
Fundador, Myriad Data

Guillermo Eldin  
Fundador, XXII

Jean-Philippe Clair

Stephanie Flacher



Cofundador de Logion Network

Rafael Frisch  
Cofundador de HawAI.tech

Alain Garnier  
Director ejecutivo de Jamespot

Antonio Graveline  
Fundador, Disaitek

Philippe Guguen  
Presidente Sorga - Mapa Emulsión

Francisco Hachem  
Fundador, Codenekt

Rodolphe Hasselvander  
Fundador, Blue Frog Robótica

Brice Hoarau  
Fundador, Semdee

Matías Houllier  
Cofundador de Wintics

Casey Joly  
Abogado, IPso

Jean Latger  
Fundador, Oktal -SE

Frédéric Lefebvre- Naré  
Director de datos de IA, Niji

Tuyo Lemrabet  
Fundador, Everysens

Jacques Lévy- Vehel  
Presidente de Case Law Analytics y  
Director de Investigación del INRIA

Olivier Laborde  
Líder de Innovación y Transformación  
Digital, BPCE

Alexandre Leforestier  
Fundador, Panodyyssey

Sixtina Lorphelin

Ingeniero de inteligencia artificial, UTC,  
estudiante en INSEAD

Aymeric Masurelle  
Cofundador de Spoon.ai

Pierre Miralles  
Cofundador de Footovision

Emmanuel Moyrand  
Cofundador de France Meta

Clément Merville  
Fundador, Manzalab - Teemew

Eduardo de Miollis  
Fundador, Polycube

Nathalie Nevejans  
Profesor de derecho privado e inteligencia  
artificial, Universidad de Artois

Rémy Ozcan  
Cofundador de Crypto4All

Jean Jacques Quisquater  
Profesor de Criptografía en la Politécnica  
de Lovaina, columnista de Le Monde

Jean-Michel Salomon  
Presidente de la Sociedad de Autores de  
Juegos

Frédéric Soufflet  
Cofundador de Haapie

Clemente Tequi  
Cofundador de Terno

François-Xavier Thoorens  
Director ejecutivo de Vaultys

Arnaud Touati  
Abogado asociado, #Hashtag

Cristophe Tricot  
Cofundador de La Forge

Killian Vermersch  
Cofundador y director ejecutivo de  
Golem.ai

# Controv3rse 解決方案

透過生成式人工智慧提昇文化和人力資本的經濟模式



Adobe Firefly

C O N T R O V **E** R S E

概括

Controv3rse 是一個獨立智庫，由 70 名企業家和生成式人工智慧專家組成。本報告針對 OpenAI 創辦人 Sam Altman 的言論給出了解決方案，面對生成式人工智慧的崛起，Sam Altman 主張建立一種獎勵創造者的新經濟模式。

必須指出的是，作品被掠奪的作者和需要明確性和法律穩定性的生成性人工智慧的開發者之間的衝突至少在 10 年內不會在法庭上解決。

同時，生成性人工智慧的激增削弱了原創性，抹去了原始作品的痕跡。同時，微任務平台正在回應對資料的大量需求，而這損害了基於創造力和版權的模型，例如 SACEM。文化資本和人力資本曾經是歐洲策略的核心，但現在正面臨雙重威脅。

然而，諾貝爾經濟學獎得主讓·梯若爾強調，國家財富的基礎是促進創新，而智慧財產權是創造力的槓桿。為了在刺激創新的同時保護文化和人力資本，新的經濟模式必須重塑作家協會。

1953 年，勞埃德·沙普利 (Lloyd Shapley) 提出了一個關鍵概念，即根據每個玩家的貢獻來公平分配合作遊戲中的獎金。這種衍生方法非常適合用於獎勵生成式人工智慧的作者，但不幸的是，其計算成本過高。還有其他替代方案，例如圖片庫使用的「付費培訓」模式 (Shutterstock、Getty)。他們根據提供的內容的數量來補償作者，但忽略了作品的品質或實用性。新創公司和研究人員正在補充該模型，以提供基於相似性的更公平的分配。我們為讀者提供了一個線上模擬器，讓讀者能夠親眼看到這種方法在實際中的技術和經濟可行性。

從長遠看，這種模式既保護了作者的權益，又反對「Uber化AI」。「薩塞姆化」重視“流動財產”，即創造力的知識財富，它在面臨壟斷企業的榨取時，能夠維持和發展文化和人力資本。

因此，我們向 Sam Altman 和生成式人工智慧的經營者推薦一種與作者協會合作的經濟模型，並向所有利害關係人推薦一種在這個新的反思框架中啟動談判的方法。

挑戰

# 1——誰希望生成式人工智慧來向作者付費？

「我認為我們需要一項新的協議、新的標準、新的協議，無論你想怎麼稱呼它，它都可以用來獎勵創造者，」Open AI 的創始人薩姆·奧特曼說。「我們需要新的經濟模式，讓創作者能夠創造收入。」

幾個月來，其他一些人士一直在倡導透過gen-AI向作者付費的想法。

## 想要向作者付費的 AI 營運商

姓名	公司	引用
薩姆·奧特曼	開放人工智慧	「我認為我們需要一項新的協議、新的標準、新的協議，無論你想怎麼稱呼它，來確定如何獎勵創作者。我們需要新的經濟模式，讓創作者能創造收入。」
阿瑟·門施	米斯特拉爾	「我們有興趣獲得優質內容。因此，我們已經開始與內容提供者、出版業和新聞界進行討論。可以找到協同效應和價值共享機制。」
大衛·霍爾茨	旅途中	「實際上，沒有辦法拍攝一億張照片並找出它們的來源。如果圖像中嵌入了一些關於版權所有者或其他資訊的元數據，那就太好了。但這是不可能的，因為沒有暫存器。沒有辦法在網路上找到一張圖片，然後自動追蹤到它的主人，然後採取任何措施來驗證它的真實性。」
克里斯·迪克森	a16z	「Web 3 運動有可能重現早期網路的精神，保障創作者的財產權，重新獲得使用者所有權和控制權，並打破大企業對我們生活的束縛。」
尼爾·史蒂芬森	Laminal (元宇宙的發明者)	「您實際上可以追蹤給定影像或輸入對最終結果的影響有多大。然後，如果最終結果有價值，那麼也許有辦法對這些投入給予讚揚和相應的補償 [...] 假設我正在寫一本書，裡面有一把書中剛剛描述的魔法劍。[...] 然後有人可以創造一個可以在虛幻資產商店出售的資產[...] 你擁有聲音設計師，他們必須對它在使用時發出的聲音做一些類似的事情。程式設計師必須使用藍圖或 C++ 將劍融入遊戲中，以便它能夠真正發揮作用並為體驗做出貢獻。因此，每一步都會增加更多的價值。在這個過程結束時，你會得到一些可以真正產生收入的東西。當收入最終到達時，您希望它向後傳播。你希望對價值鏈做出貢獻的不同人都能得到某種形式的補償」
傑倫·蘭尼爾	微軟	「人們將因其數據而獲得報酬，並將為需要其他人數據的服務付費。個人的注意力將由他們自我定義的興趣引導，而不是被廣告商或其他第三方操縱的平台所引導。平台將獲得更高品質的數據來訓練他們的機器學習系統。」拉尼爾承認，即使是資料尊嚴研究人員也無法就如何解開人工智慧模型所吸收的一切或需要多少細節來解釋這一切達成一致。然而拉尼爾認為這是可以實現的——只是逐步地。「我們需要計算並呈現對給定的人工智慧結果最重要的人類來源的來源。我們目前不這樣做。我們可以有效且有效率地做到這一點，但我們卻沒有這樣做。採取這一舉措必須由社會決定。」
蕭逸	Animoca品牌	「數據是最寶貴的資源。首先，它們為我們經常聽到的所有人工智慧提供動力。沒有數據，就沒有ChatGPT。數據就是新的工作。我們並沒有得到公平的補償。[...] 產權和國家財富之間存在著明顯的相關性。」想像一下，你登入Facebook後一天結束時意識到你剛剛為公司賺了一千美元。您與Facebook的關係將徹底改變。您或許會要求您獲得應得的份額。」



姓名	公司	引用
楊樂村	元	“一種向一個方向傳播版權、向另一個方向增加收入的模式”
特里普·阿德勒	Scribd 前首席執行官	「第四定律是一套指導人工智慧公司如何使用和訓練人類生成內容的原則。第四定律受科幻小說作家艾薩克·阿西莫夫的《機器人三定律》啟發，規定人類應該有權同意和控制人工智慧如何使用其工作，並獲得報酬（如果提出要求）和對其工作的認可（如果一本書被結果引用，應該有一個購買連結）。」

在 Sam Altman 的領導下，對作者進行補償的想法變得更加實質性，儘管他認為這種補償只有在作者的風格被抄襲時才會產生，這非常具有局限性。

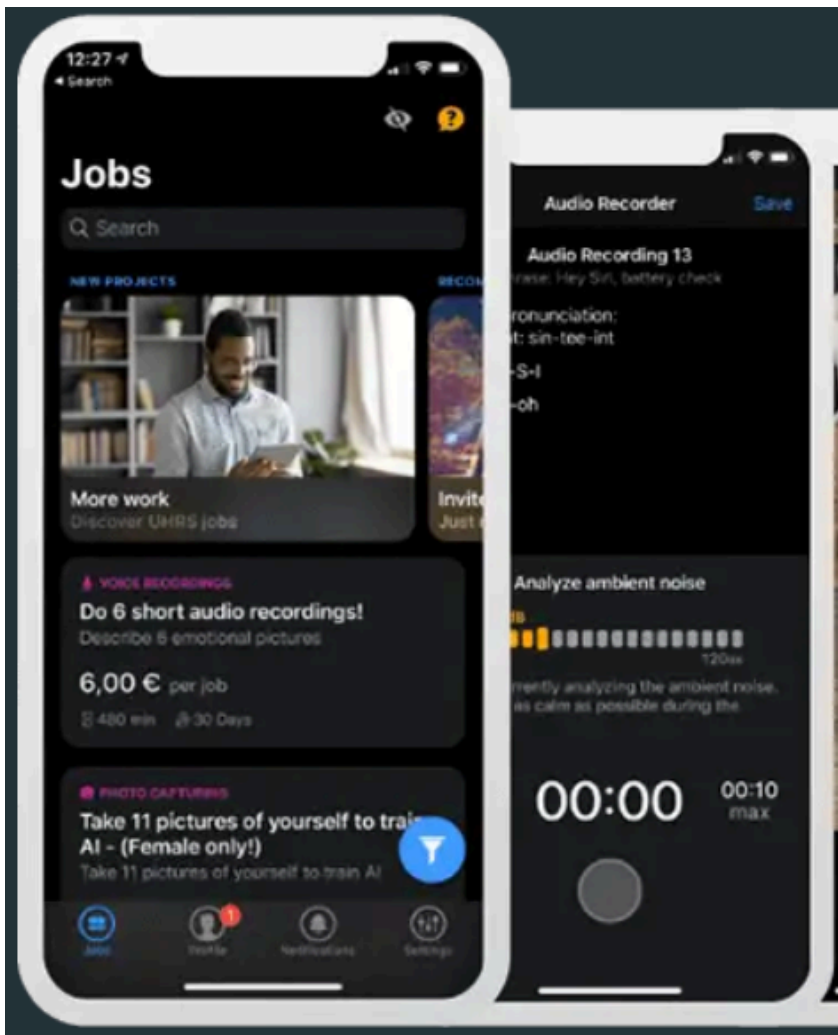
實施的經濟模式提出質疑。本說明的目的是在人工智慧法案的框架內提出這種新的經濟模式。

## 2 – 作者報酬的三個方面

如果單純從法律層面處理，作者與生成人工智慧業者之間的抓取之爭將持續至少十年。

同時，產業將細分為數據創造者、平台、聚合者、談判者、培訓者、提煉者、開發者和生成人工智慧的整合者，這將淡化創造行為。文化遺產將被混合、傳播和循環利用，其原有的痕跡將被抹去。

由於擔心學習內容退化，一個新的微任務領域已經出現，用於大規模生產內容。



©Clickworker 2024

培育文化資本。

在其他經濟模式中（見下文），資料抓取、資料優步和作家協會是作者獲得報酬（或不獲得報酬）的三大主要方面，這種不平衡加劇了當前的緊張局勢。

零工經濟平台是基於命令邏輯。Clickworker直接向用戶支付 6 歐元，讓他們大聲描述 6 張圖片並發布錄音。或拍一張自己鍛鍊身體的照片。Gigwalk和Premise支付 4 美元來拍攝路口交通擁堵的照片。此結果用於訓練人工智慧。

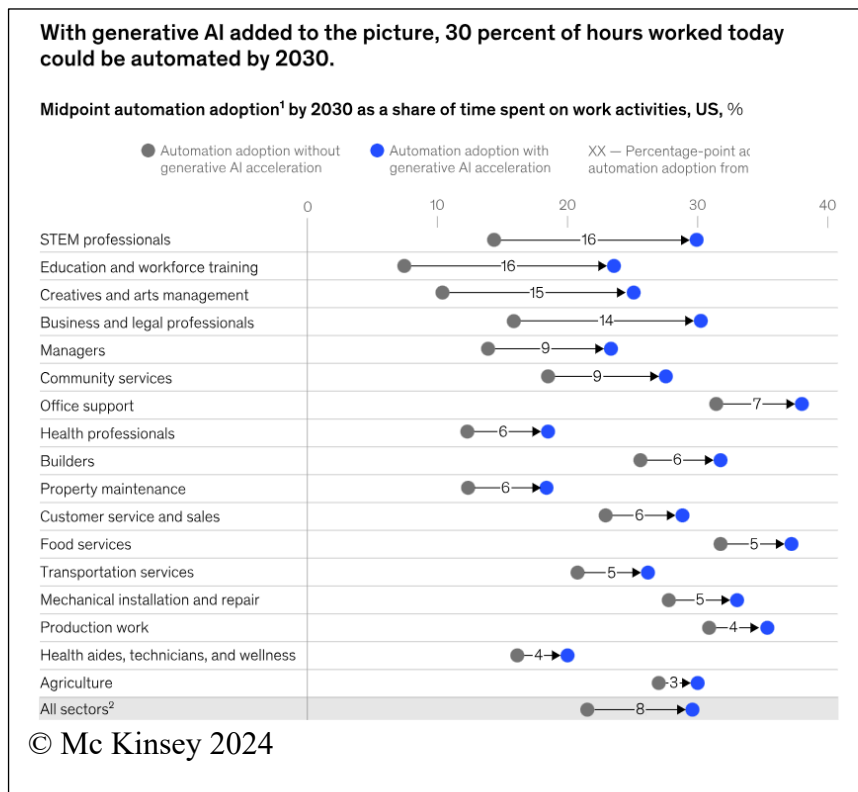
由於生成式人工智慧營運商對數據的大量需求，這種「數據優步」類型的模型正在迅速發展。

這與 SACEM 等版權協會相反，這些協會以作者的供應和智慧財產權為基礎，追蹤作者的財產、收取版稅，並根據預先定義的分配比例向權利人支付報酬。

第一種模型偏向需求、執行、直接商品化、認知新泰勒主義、工作的標準化。第二促進供給、主動性、智慧財產權、創造力、文化多樣性。前者開發文化資本，後者



### 3 – 除了版權，人力資本



面對生成式人工智慧的發展，作家和創造力處於最前線。接下來是所有經濟部門的手工藝、藝術專業、專業知識、研究和所有具有專門技術的職業。

隨著「數據的Uber」模式越來越普及，它將催生一種Uber化的人工智慧，就像Uber公司曾經開創的Uber化現象一樣。

例如，透過客製化企業生成式人工智慧及其所有資料檔案，您可以獲得每個新訊息、報告、專案、簡報和定價的草稿。專業的糾錯會對AI進行訓練，直到錯誤率變得可以接受。生產力的提高是顯著的，但作為回報，企業在價值創造策略環節上的地位卻在下降，這對人工智慧營運商來說是有利的。

由於所有這些技能都是知識、實踐和社會組織材料的交匯，因此，企業的結構處於危險之中。這就是股市對大型科技公司的預期。透過授予數萬億美元的資本，它不再看重公司，而是重視壟斷。

機器擁有免費「學習權利」的觀念將會傳播開來。十年後，即使作者最終贏得了針對數據抓取的法律鬥爭，他們的報酬也將基於Uber提供的數據可比性，也就是說，報酬不會太多。回過頭來看，我們會發現現有的版權不過是一條可以輕易繞過的馬其諾防線。

歐洲經濟擁有文化資本和人力資本，長期以來被視為寶貴的資產，甚至被置於歐洲經濟戰略的中心。這個過早的策略最終因缺乏經濟模型而被遺忘。由於價值被低估且保護不力，這片資本現在面臨著被Uber化的人工智慧奪走的威脅。對每個公司、國家或大陸來說，拒絕科技意味著走向持久的衰落，就像中國兩個世紀以來的衰落一樣。不加補償地接受它，就等於不經鬥爭而被殖民。

諾貝爾獎得主讓·梯若爾教導我們：1/ 國家的財富取決於其在創新層面獲取價值的能力。2/ 創新源自於個人的主動性，因為創造性勞動的特徵就是我們不知道我們將會發現什麼。3/ 智慧財產權是激發這種創造力的必要之惡。4/ 阻止一家公司行銷其產品直到它支付版稅是針對Gafam非常有效的武器。5/ 具有單獨許可授權的智慧財產權池將要求他們以有競爭力的價格獲得許可。

想要增強人力和文化資本的國家必須找到現代化的科技解決方案，其經濟模式將有效地與Uberizing AI 競爭。

# 技術方案

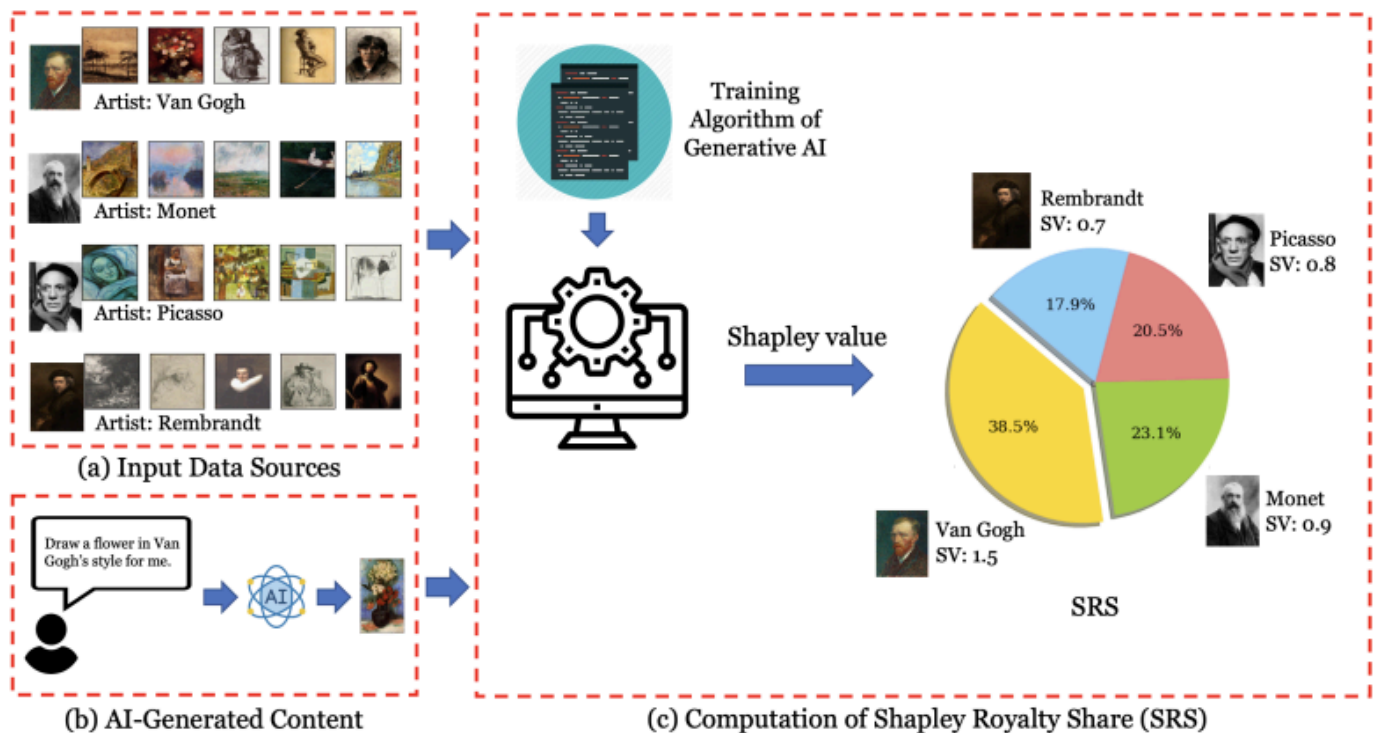
## 4-理想的解決方案

在博弈論中，勞埃德·沙普利 (Lloyd Shapley) 於 1953 年提出了這樣一種觀點，即在合作遊戲中，可以計算出玩家之間公平的收益分配。例如，我們考慮一個包含所有生物物種（藻類、海葵、貝殼、海星）的海洋生態系統，並根據每個物種對生態系統生命的貢獻來衡量其用處。當你移除海星時，50%的其他物種會消失，而當你移除海葵時，只有 10%的其他物種會消失。海星的沙普利值是海葵的五倍。

研究人員對於作者報酬的理想解決方案已經達成共識。這涉及計算每個作者的 Shapley 值，並根據該值按比例分配報酬。

對於生成性人工智慧而言，訓練作品的報酬公平分配最好由作者根據訓練作品對人工智慧生成內容的有用性來進行。下圖中，梵谷的 Shapley 值為 1.5，即總補償金的 38.5%。

Shapley 值



在實務中，計算 Shapley 值的成本過高，引發了公平報酬解決方案的技術可行性問題。

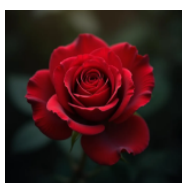
## 5 – 演示者

### 目的

讀者可以在[https://huggingface.co/spaces/TheFrenchDemos/genai-rew\\_a\\_rd](https://huggingface.co/spaces/TheFrenchDemos/genai-rew_a_rd)找到線上模擬器








我們相信，現在從技術和經濟角度來看，都可以計算出培訓內容的歸因分數，以便在該內容的版權持有者之間公平分配報酬。這個模擬器將允許每個人自己驗證自己的經濟假設。

盛開的鮮豔紅玫瑰，「微距照片」產生一張玫瑰影像：



模擬器在幾秒鐘內從 Open Images 資料庫的一千萬張內容中找出對生成內容的形成貢獻最大的八張原始圖像。它計算歸因率、補償，並提供有關原始內容及其作者的身份資訊。

### 玫瑰影像的屬性

內容	歸因	內容獎勵 (c€)	內容網址	作者姓名	作者網址
	21%	0.63	<a href="https://www.flickr.com/photos/50093642@N03/4629227563">https://www.flickr.com/photos/50093642@N03/4629227563</a>	Yoko Nekonomania	<a href="https://www.flickr.com/photos/nekonomania/">https://www.flickr.com/photos/nekonomania/</a>
	18%	0.54	<a href="https://www.flickr.com/photos/22305657@N02/11733854773tt">https://www.flickr.com/photos/22305657@N02/11733854773tt</a>	布魯諾皮尼奧	<a href="https://www.flickr.com/photos/brunopinho/">https://www.flickr.com/photos/brunopinho/</a>
	15%	0.45	<a href="https://www.flickr.com/photos/12692992@N00/170278141">https://www.flickr.com/photos/12692992@N00/170278141</a>	迪恩 + 巴布	<a href="https://www.flickr.com/photos/barbandean/">https://www.flickr.com/photos/barbandean/</a>
	12%	0.36	<a href="https://www.flickr.com/photos/33140999@N07/3509399376">https://www.flickr.com/photos/33140999@N07/3509399376</a>	米歇爾·阿里吉	<a href="https://www.flickr.com/photos/arrighi/">https://www.flickr.com/photos/arrighi/</a>
	10%	0.30	<a href="https://www.flickr.com/photos/55658968@N00/8730223835">https://www.flickr.com/photos/55658968@N00/8730223835</a>	蒲生俊博	<a href="https://www.flickr.com/photos/dakiny/">https://www.flickr.com/photos/dakiny/</a>
	9%	0.27	<a href="https://www.flickr.com/photos/34226407@N04/8734050294">https://www.flickr.com/photos/34226407@N04/8734050294</a>	艾德·普林斯	<a href="https://www.flickr.com/photos/edpics/">https://www.flickr.com/photos/edpics/</a>
	9%	0.27	<a href="https://www.flickr.com/photos/41353201@N07/4267426012">https://www.flickr.com/photos/41353201@N07/4267426012</a>	音樂家洛克斯頓	<a href="https://www.flickr.com/photos/dj-dwayne/">https://www.flickr.com/photos/dj-dwayne/</a>

	8%	0.24	<a href="https://www.flickr.com/photos/64574638@N04/6006728032">https://www.flickr.com/photos/64574638@N04/6006728032</a>	塞爾維亞人	<a href="https://www.flickr.com/photos/serbestruhpro/">https://www.flickr.com/photos/serbestruhpro/</a>
---	----	------	---	-------	---

在網頁底部，有三個按鈕允許使用者下載：

- 本報告
- 運營商可以產生權利人報告，並在適用和解協議和公平報酬返還時傳達給權利人
- 按照文學藝術最高委員會提議的形式，由AI法案（第53條，1，d）規定的法定報告物業（CSPLA）

## 工作原理

模擬器有一個“高級”按鈕 參數»允許使用者調整他們想要的參數。下表複製了預設設定並指明了計算公式。

## 模擬的參數和值

多變的	環境	可比	計算	公式
使用者訂閱	12 歐元/月 (1)	Adobe Firefly起價 4.99 美元/月 中途旅程每月 10 美元起 Dall-e包含在ChatGPT Plus 中 \$20 Getty Edify 25 代 45 歐元		年代
產生內容	60 /月	Adobe Firefly基本方案 100/月 Midjourney基本計劃 100/月		格
每份產生內容的面值			0.2 歐元	勝利者=勝率/敗率
權利人和作者的份額				$R\% = Rr\% + Ra\%$
- 權利人	10%	音樂 (ASCAP、BMI、SACEM) 2 – 8% 庫存圖片 55 – 85%		相對風險%
- 作者	5%	紙本書 5 – 15% 音樂（表演者和作曲家） 15 – 30% 庫存攝影 15 – 45%		鉅%
版權持有者和作者按生成內容所獲得的報酬			3 歐元	$R = V \times R\%$

歸因		21%	A% = 相似度演算法 N = 分配的內容數量
		8	
培訓內容獎勵		0.63 歐元	T = R x A%
作者人數	1,800		乙
使用者	50萬	Midjourney的付費訂戶數量估計在200 萬至 500 萬之間。	烏
權利人和作者的月度獎勵		829 歐元	$S \times U \times R\% / B \times A\% \times N$
- 權利人		276 歐元	$S \times U \times Rr\% / B \times A\% \times N$
- 作者		553 歐元	$S \times U \times Ra\% / B \times A\% \times N$

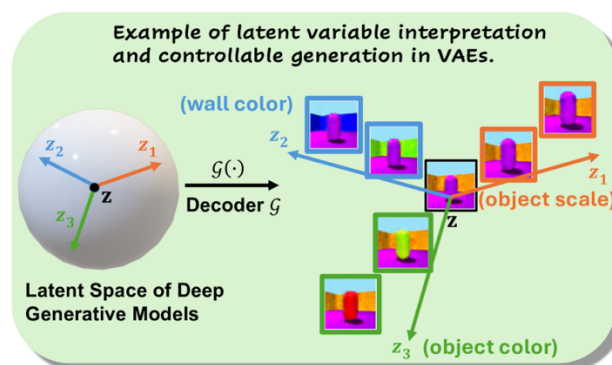
## 相似性的概念

此演示器使用訓練內容和生成內容之間的相似度演算法，實現了[DINOv2](#)和[FAISS](#)。直觀地看，兩張相似的圖像看起來很相似。

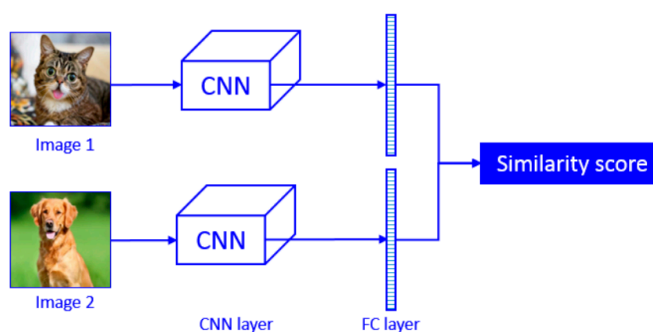
但這裡所說的「相似」是一個科學概念，與「假冒」、「相似」或侵權等法律概念無關，法國法律將其定義為「以任何手段進行的任何複製、再現或分發」。作者權利的行為，均視為侵犯了法律規定和規範的智力作品」（CPI，第335-3條）。複製作品的某個元素（在本例中為「貢獻」）不一定構成侵權或侵權。同樣，當作者創作時，一個元素可以進入其作品中，而不會存在任何第三方的偽造或侵權。因此，因「相似」而產生的報酬不應該在侵權的情況下阻礙作者採取行動。

## 向量相似度

從技術上講，每個生成或原始內容都會經過分析，以提取可測量的資訊：貓、狗、前景色和大小、類型、樣式、大小、識別的物件、關鍵字...所有這些數百個實用指標，在我們的演示器中大約有500個，在「潛在空間」中形成一個向量。



作品（圖像 1）對生成內容（圖像 2）的歸屬程度是透過兩個暹羅神經網路處理兩個圖像之間的相似度得分來計算。得到的向量越接近，影像之間的相似度分數就越高。



## 作者人數

Open Images 資料庫是由 Flickr 資料庫改進而來，採用了諸如 Google 影像搜尋不偵測、最小像素數或重複資料刪除等篩選器。Open Images 也只選擇了 CC-BY 許可證，每 30 張圖片中就有一張是採用該許可證的。這表明最活躍的貢獻者在 Open Images 中的集中度很高，他們發布了超過 5,623 張照片，或者說大約 1,800 位作者。這個直覺的假設可以透過對 Open Images 資料庫的作者進行精確計數來驗證，但由於查詢速度的限制，我們無法實現這一點。

## 作者身份確認

Open Images 資料庫可辨識圖像的作者，並提供他們的 Flickr 個人資料頁面，但其中並不包含完整的電子郵件或郵寄地址。

除了我們的演示器和開放圖像之外，公共資料集還提供了精度可變的識別元素，但可以交叉引用以識別每個權利持有者。



交叉引用方法最廣為人知的是其可欺詐性地侵犯隱私。早在 1997 年，馬薩諸塞州一家保險公司強制報告匯總統計數據以及選舉數據，就可以毫無差錯地識別州長威廉·韋爾德( William Weld ) 的身分並了解他的健康記錄。

在大多數情況下，合法使用可以完全識別作者：透過比較社交圖譜、透過罕見的屬性組合（科學領域、活動時間、IP 位址）進行統計相關性、透過重建個人屬性來自群體平均值和其他個體的已知屬性等下表提供了識別 GPT 4 訓練內容作者的線索。



AI-generated by Adobe Firefly

#### GPT-4 所使用的資料集作者的潛在身分識別

	數據集	識別要素	直接交叉	識別符
直接識別	打開圖片	-	-	姓名、姓名、個人資料表
	古騰堡計劃	-	-	姓名、名字、出生和死亡日期
	論文集	-	-	姓名，首字母
	PubMed	-	-	id_author
間接識別	常見爬蟲	領域	Whois	Registrant_contact (姓名、組織、街道、城市、州、電話)
	書籍語料庫	純文字	粉碎詞	姓名，名字
	網頁文本	Plain_text + 搜尋 = “作者”	亞馬遜 API	姓名，名字
	StackOverflow	HTML	Stackoverflow.com	使用者 ID、網路設定檔
在 等待索賠	維基百科	使用者名稱	-	-
	Reddit	使用者名稱	-	-
其他	共同之聲	(統計數據)	-	-
	Google 圖書Ngrams	(不適用)	(不適用)	(不適用)
	GitHub	(無法訪問)		

## 6 – 現有解決方案

### 合成

下表列出了與我們的演示器相關的一些商業化或測試過的方法：

#### 推廣作者內容的方法

估價	估值方法	示威者	商業公司	研究者
產生內容	訂閱費用/數量	爭議		
數據集	內容集的相似性 沙普利 運行中數據 Shapley 網站存取頻率、資料類型		收費站	Adobe 開放人工智慧 開放人工智慧
培訓工作	內容相似性 建模學習 模板定制 忘記 刺青 Shapley 聯合	爭議	Bria (圖片)、Prorata Gist.ai (文)	開放人工智慧 Adobe Adobe Adobe 華為
作者	付費培訓 作品整體影響力 權利人身份識別		Shutterstock、Getty、 Audiosparx、Canva Sureel.ai (音樂)	
		爭議		

## 已上市的解決方案

### 向培訓內容作者付費的公司

公司	創投	週轉	方法	媒體	內容提供者
Adobe (螢火蟲)	-	210億美元	付費培訓	圖片	
燭光	-	20億美元	付費培訓	圖片	
蓋蒂	-	9.17億美元	付費培訓	圖片	
Shutterstock	-	8.75億美元	付費培訓	圖片	
ProRata (Gist.ai)	2500萬美元	-	內容相似性	文字	天空新聞台、時代雜誌、金融時報、大西洋月刊、Axel Springer、衛報、dmg media、環球音樂、財富
布里亞	2400萬美元	-	內容相似性	圖片	Shutterstock、Getty Images、Alamy、Superstock
收費站	2400萬美元	-	造訪網站的頻率	文字	Penske Media Corporation、時代、Mumsnet、Trusted Media Brands、Candr Media Group、AdWeek
蘇瑞爾	-	-	作品整體影響力	影像、音樂	
音訊	-	-	付費培訓	音樂	

### 付費培訓方式

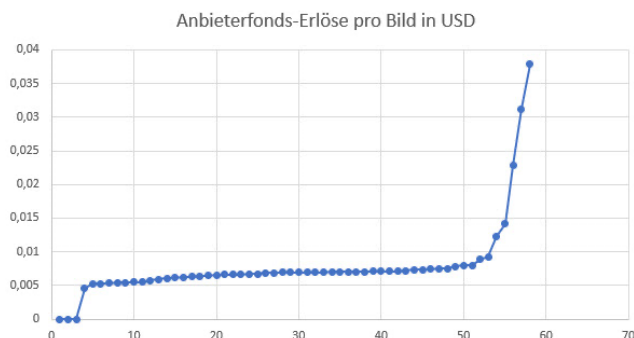


AI-generated by DALL-e

Adobe 的私人內容庫Canva、Getty 和 Shutterstock採用「付費訓練」方式分配其出售的資料集的價格。作者根據他們對這些數據集的貢獻量獲得報酬。

「付費培訓」的價值可以透過各種效用指標來計算。例如，Shutterstock每張照片每年支付 0.0050 美元，最高可達每張照片每年 0.0038 美元，具體取決於元資料的數量和其他指標。

付費培訓價值： Shutterstock作者的報酬。樣本N = 58（美元/張圖片）



此模型的優點是計算簡單。這個系統已經運作兩年了，似乎已經得到了權利人的廣泛接受。

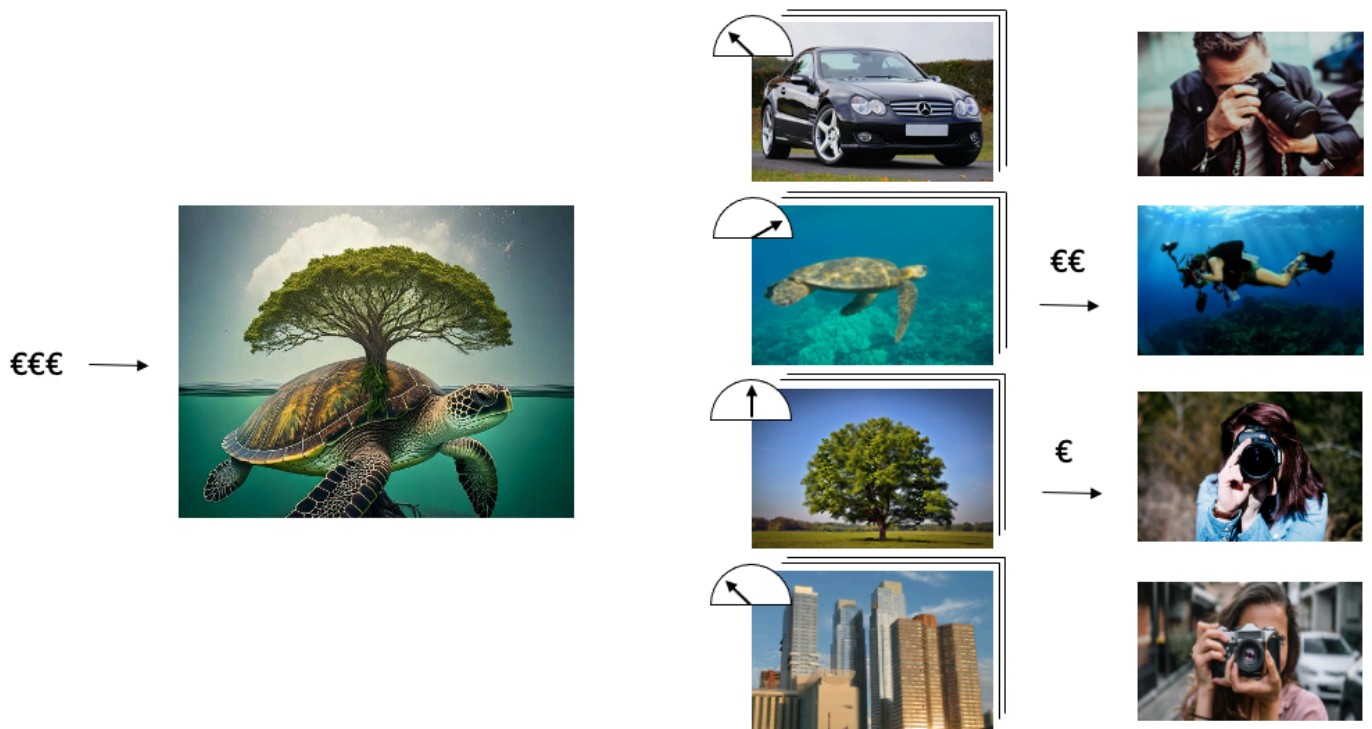
另一方面，作者之間的公平程度呈現兩種偏見：

- 內容的品質不予考慮：在其他條件相同的情況下，貢獻 200 張照片的作者獲得的報酬是貢獻 100 張照片的作者的兩倍。
- 生成內容的實用性不予考慮：如果用戶在提示中更頻繁地插入“以梵高的風格”，梵高所獲得的評價並不會比畢卡索多。

## 混合方法 相似度+付費訓練

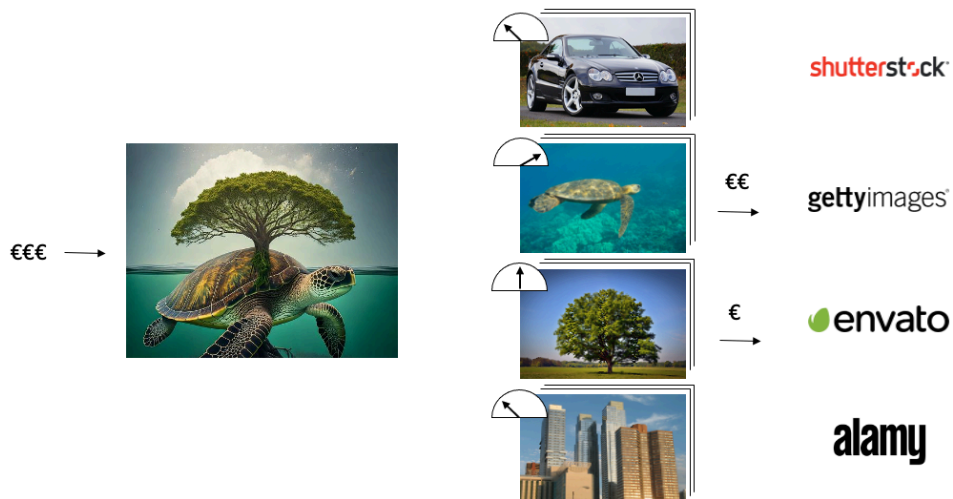
以色列-美國新創公司Bria 與大型私人圖像庫（Shutterstock、Getty）合作，並被大型科技公司（微軟、亞馬遜、Nvidia）整合。作為第一種方法，Bria 與我們的演示者一樣，使用相似性方法，以便根據所測量的相似性按比例向作者支付報酬。

衡量生成內容與原始照片之間的相似度，以獎勵作者



更準確地說，相似度得分的總和似乎是由圖像庫（Shutterstock、Getty、Alamy .....）計算的，這會導致銀行獲得與這些總和成比例的初始報酬。

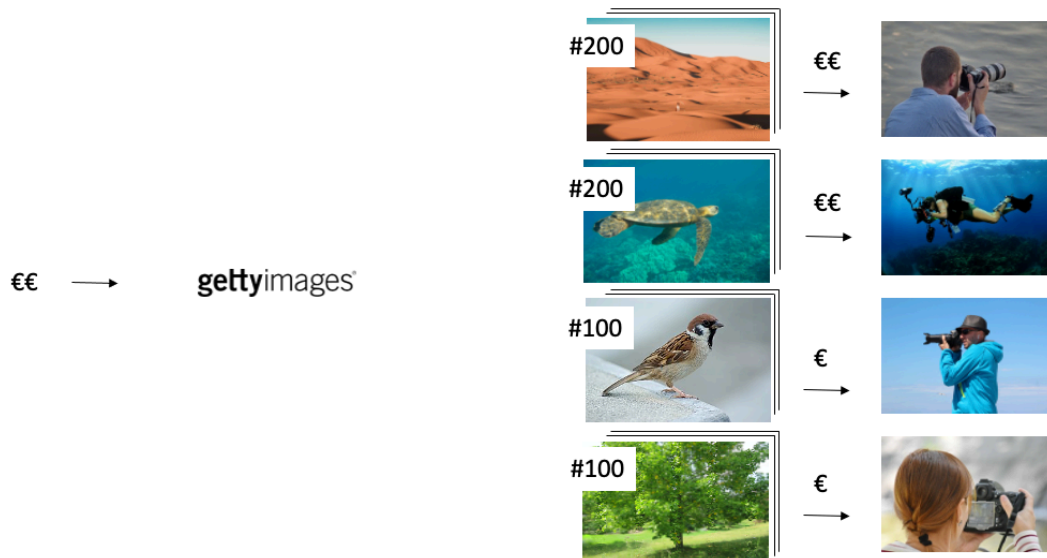
測量生成內容與原始照片之間的相似度，以獎勵圖片庫



付費訓練」的方式，根據作者對所售資料集的貢獻作品數量，在作者之間分配第一筆報酬。



## 向作者支付照片數量



經過進一步的改進，具有高品質元資料的影像將獲得獎勵。

因此，Bria 的方法分為兩個階段：

- 首先，得益於訓練影像和生成影像之間的相似性演算法
- 然後，得益於作者對所售資料集的量化貢獻演算法（「付費訓練」）

這種混合解決方案是：

- 經濟上與生成式人工智慧的商業模式相容
- 被圖片庫和作者接受兩年
- 目前正在擴展到其他類型的內容：3D、音樂、語音、動畫圖像、視頻

提的是，新創公司ProRata很可能採用了這種混合方法，並將其提供給已經與 AI 營運商達成協議的媒體。它涉及「獲取生成式人工智慧的輸出，無論是文字、圖像、音樂還是電影，並將其分解成各個組成部分，找出它們的來源，然後向每個版權持有者分配一定比例，並向他們支付因此。其母公司的公開專利並沒有提供進一步的細節。

這種混合方法仍然存在一些缺陷：

- 尚未證明該解決方案是否適用於所有類型的內容，例如遊戲
- 無法保證相似性不是巧合。換句話說，即使作者的作品對於生成的內容沒有用處，他仍然可以獲得報酬，反之亦然。
- 對原始影像的貢獻度進行近似測量。在上面的例子中，沙漠攝影師與海龜攝影師的報酬相同，因為他們對Getty Images資料集的貢獻數量相同，而沙漠照片對產生的內容沒有任何用處。

需要注意的是：

- 即使對特定生成內容的補償分配是近似的，其公平性也可能隨著生成內容數量的增加而提高。
- Bria 無需對作者進行單獨識別，只需識別訓練資料集即可

- 分發演算法本身並不公開。作者將來可能會要求更多的透明度，但也有可能相反，需要一定程度的保密性以避免機會主義偏見。
- 我們也不知道作者相對於我們認為可以出版的內容的營業額的平均報酬率。

## 資料集相似性方法

新創公司Sureel為音樂家提供了一種工具，可以測量他們的作品對 AI 生成的曲目的影響力（以百分比來衡量）。

他在 '949 專利中描述了這種演算法，該演算法根據作者所有作品的影響力來預測作者的報酬，例如透過基礎模型的權重與該集合對其進行微調的權重之間的差異來衡量。這種變化旨在透過比較更多數量的作品來淡化相似性偏見。

## 目前正在測試的方法

除了上述幾種變化之外，研究人員還在競相嘗試新方法：

### 尋找公平的薪酬模式

商業	方法	研究團隊	日期	動機
開放人工智慧 普林斯頓大學 哥倫比亞大學 哈佛大學 賓州大學 加州大學 柏克萊	沙普利	生成式人工智慧版權挑戰的經濟解決方案 Jiachen T. Wang, 鄧準, Hiroaki Chiba-Okabe, Boaz Barak ,蘇偉傑	2024 年 4 月 14 日 2024 年 9 月 9 日 (v4)	Shapley值提供了一種基於效用分配收益的原則性方法。 此實用程式有助於衡量資料來源對生成內容的負責程度。如果反事實模型不太可能產生與部署模型相同的內容，則該值較低，反之亦然。 效用可以解釋為所有成員因提供資料來訓練生成式 AI 模型而應得的總補償。
維吉尼亞理工大學 麻省理工學院	運行中數據 Shapley	一次訓練運行中的資料 Shapley Jiachen T. Wang, Prateek Mittal、Dawn Song 和Ruoxi Jia	2024 年 6 月 29 日	超級對齊」) 教授、同時在OpenAI工作的Boaz Barak並沒有忽視 Shapley 的理想方法，即使這意味著在準確性上做出妥協。 為從類似 GPT 的資料集訓練出來的通用 AI 模型計算 Wikipedia 或Github等資料集的 Shapley 值。 團隊表示，將報酬與每個資料集的量化貢獻相結合，可以保證版稅分配的最大公平性。她表示她更喜歡成本較低的運行內數據 Shapley 變體。



商業	方法	研究團隊	日期	動機
	訓練建模	<b>TRAK : 大規模歸因模型行為</b> 樸成敏、Kristian Georgiev、Andrew Ilyas、Guillaume Leclerc、Aleksander Madry MIT	2023 年 4 月 3 日	亞歷山大·馬德里 (Aleksander Madry) 在 OpenAI 負責人工智慧推理。Trak 解決方案依賴 AI 模型的近似表示來描述其在訓練過程中的演變，並將預測與影響它們的訓練資料聯繫起來。  事實證明，該解決方案對於在 ImageNet、視覺語言模型 (CLIP) 和語言模型 (BERT 和 mT5) 上訓練的影像分類器非常有用。
Adobe 卡內基美隆大學 加州大學柏克萊分校	數據集相似性	<b>評估文字轉圖像模型的資料屬性</b> 王勝宇, Alexei Efros, 朱俊彥, Richard Zhan	2023 年 8 月 8 日	由於相似性方法很可能揭示巧合的相似之處，因此可以透過將此方法擴展到兩組圖像 (訓練和生成) 來減弱這種偏見。Adobe 研究人員由此計算出第一組資料集對一組合成內容的貢獻
	水印	<b>ProMark : 用於因果歸因的主動擴散浮水印</b> Vishal Asnani、John Colomosse 等人	2024 年 3 月 14 日	紋身展示了訓練內容和合成內容之間的因果關係。這使得對創作者的合法歸屬或付款可以有信心地建立。
	忘記	<b>透過取消學習合成圖像實現文字到圖像模型的資料歸因</b> Sheng-Yu Wang、Aaron Hertzmann、Alexei A. Efros、Jun-Yan Zhu、Richard Zhang	2024 年 6 月	此方法透過代理訓練被遺忘的影像來取消合成影像的學習，在取消學習過程後識別出那些具有顯著損失偏差的影像，並將其標記為有影響力的影像。
華為 不列顛哥倫比亞大學 西門菲沙大學 瑞典皇家理工學院	聯邦 Shapley	<b>提高橫向聯邦學習中數據估值的公平性</b> 範哲安, 方黃, 週子瑞, 裴健, Michael P. Friedlander, 劉長新, 張勇	2022 年 5 月 23 日	數據所有者的動機在一定程度上取決於聯邦學習中協作和獎勵的公平性。

## 變體和發展

在所有提到的方法中：

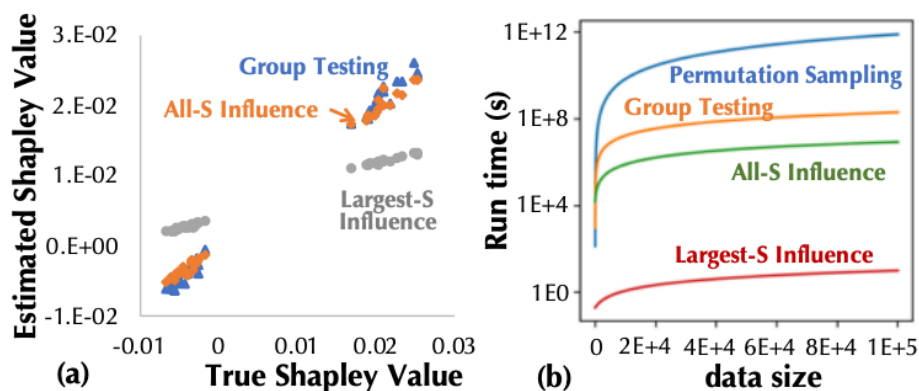
- 報酬參數可根據實用性指標的仲裁進行調整：大小、日期、相似度、元資料數量、獨立商業成功等。

### 指標 實用性

公司研究員	媒體	效用指標
樸晨等	圖片	向量
王,鄧等.	百科全書、學術出版、圖像、程式碼	反事實模型 (S) x( gen ) 模型與反事實模型產生的內容
王, 米塔爾等人。	百科全書、學術出版品、圖像、程式碼、電子郵件、論壇	保留驗證資料點集 Z(val)
鄧、張等。	音樂	事件的平均速度 所有項目的平均高度 所有事件的時間差總和
Shutterstock + Bria	圖像、向量圖	相似度評分 按資料集劃分的收入 每個資料集和每個作者的資料量 元資料磁碟區 在原始模型開發中發揮的作用 透過與未來生成性許可活動掛鉤的特許權使用費發揮作用
Adobe Stock + Bria	圖像、向量圖	相似度評分 影像可接受性 圖片日期 下載次數

- 該演算法可以在精度和執行速度之間進行調整。具體來說，Shapley 值可以計算
  - 產生任意數量的內容
  - 具有任意粒度（按資料集、子資料集、直至單一資料）
  - 具有類似功能

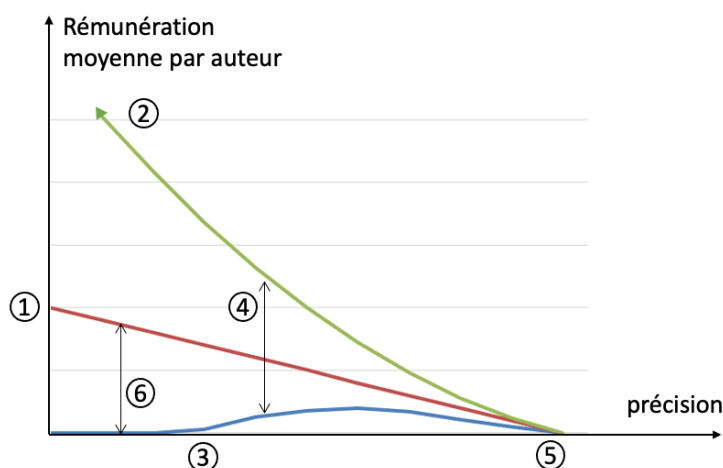
### 三個近似 Shapley 值的函數



在現階段，研究領域已經足夠廣泛，希望找到透過使用來優化公平、成本和精確度之間關係的報酬解決方案。

### 精準成本對薪酬的影響。

- (31) 在零精度下，每位作者的報酬等於總藏書量除以作者人數。
- (32) 在零精準度下，作者的最大潛在報酬等於總收藏量
- (33) 在零精度下，作者的最低潛在報酬等於零
- (34) 不確定性差距
- (35) 當精確度成本超過總收集量時，補償和不確定性差距為零
- (36) 在採集量不變的情況下，每位作者的平均報酬隨著精確度的提高而降低



## 7 – 可接受性

**JDN**  
JOURNAL DU NET

**#FairlyShare contre le péché originel d’Internet – septembre 2014**, par Vincent Lorphelin, Gilles Babinet et al.

Ce qu’ils dénoncent ? Les plate-formes qui reposent sur l’industrialisation du travail gratuit, en dehors des règles de l’appel d’offre, du concours ou du bénévolat. Ils emboîtent ainsi le pas à [...] la levée de boucliers qui a suivi le rachat du journal de crowdsourcing Huffington Post par AOL pour 315 millions de dollars. [Lire](#)

付費訓練」方式是目前最為普遍的對使用作品訓練通用人工智慧模型的作者進行報酬的方式。儘管具有一定的透明度和公平性，但它尚未引發任何訴訟或請願，例如在「免費勞動力」宣傳或《赫芬頓郵報》出售期間所見（見方框）。

## 媒體銀行使用「付費培訓」和使用其數據集的人工智慧模型

日期	媒體	資料集提供者	人工智慧模型提供者
10/25/22	圖片、照片	Shutterstock	OpenAI
01/12/23	圖片、照片	Shutterstock	Meta、谷歌、亞馬遜、蘋果
03/12/23	圖片、照片	Shutterstock	LG (艾克森)
03/21/23	圖片、照片	蓋蒂圖片社	英偉達
05/29/23	阿凡達	崛起藝術	
07/11/23	圖片、照片	Shutterstock	OpenAI
08/08/23	圖片、照片	Shutterstock	Nvidia (畢卡索)
09/07/23	圖片、照片	蓋蒂圖片社 阿拉米 布里亞	

日期	媒體	資料集提供者	人工智慧模型提供者
09/13/23	音樂	音訊Sparx	穩定性人工智慧
09/13/23	圖片、照片	Adobe Stock (螢火蟲)	
09/26/23	圖片、照片	蓋蒂圖片社 (Getty Gen AI)	
10/04/23	圖片、照片	Canva (魔術工作室)	開放式人工智慧 (Dall-e) Google (圖)
2024年2月21日	圖片、照片	超庫存k 恩瓦托	
04/11/24	圖片、照片	眼愛 自由挑選 相簿	
06/27/24	音樂、語音	環球音樂* 索尼娛樂 華納唱片	Google (Youtube)
07/30/24	3D	Shutterstock	Nvidia (Edify)

\*實驗

資料集購買合約都附有宣傳「尊重、負責、安全和公平」的作者報酬的演講。

## 關於付費培訓公平性的演講

平台	庫存和數據集	聲明
英偉達	Getty、Shutterstock、Adobe	Nvidia：「我們創建了一個平台，讓我們的合作夥伴能夠利用授權資料進行訓練，例如 Getty、Shutterstock或 Adobe 的資料。」黃仁勳說。「他們尊重內容所有者。訓練資料來自這個來源，由此產生的任何經濟利益都可以回饋給創造者。」
	Adobe	Adobe：我們以創作者為中心，負責任地開發生成式人工智慧。 Adobe：放心創作，因為您的輸出具有最高品質，並且採用可安全用於商業用途的範本建構。
	蓋蒂	Getty：iStock的生成式人工智慧是一款經濟實惠且具有商業安全性的生成式人工智慧工具，無需擔心受法律保護的內容會滲入資料集並最終影響到他們的工作
元	Shutterstock	Shutterstock：擅長建立大型生態系統，獎勵貢獻者並連結創作者。
LG Exaone	Shutterstock	Shutterstock：人工智慧生成內容的所有權不能歸屬於個人，而必須補償參與創作每件新內容的眾多藝術家。我們的客戶可以放心、合法地授權 AI 影像以滿足自己的需求。我們也會為那些訓練我們模型的貢獻者提供報酬，因此您也可以放心地使用我們的人工智慧。我們的 AI 影像產生器仍然是您熟悉且喜愛的安全、易用且設計合理的功能  Shutterstock的數百萬張高解析度影像和元資料進行訓練，會將文字提示轉換為影像。作品被用於訓練模型的貢獻者將透過Shutterstock的貢獻者基金獲得報酬，而且每當使用其智慧財產權創作出新的生成內容並由客戶授權時，他們也將獲得報酬。
開放人工智慧	Shutterstock	OpenAI：我們從Shutterstock獲得許可的資料對於訓練 DALL-E 至關重要。確保創作者經濟持續充滿活力是OpenAI的首要任務。作家、藝術家、作曲家和其他創作者在整個文明史上為社會做出了不可估量的貢獻[...] OpenAI並不想取代創作者。
	燭光	Canva：Magic Media的文字轉圖像轉換器、OpenAI的 DALL·E和Google Cloud的Imagen [...] 確保他們的 AI 模型公平且合乎道德。「我們與多年來一直與我們在一起的創作者分享我們在Canva中開發的價值。透過新的創作者基金，他們可以獲得與 AI 創作相符的版稅池。」
Google雲	燭光 Shutterstock	Google Cloud：Shutterstock也率先推出了符合道德標準的 AI 圖像生成器，成為了領先的創新者，目前該生成器已透過Vertex AI 上的Imagen得到增強。
谷歌Youtube	環球音樂集團	谷歌子公司 YouTube 正嘗試與音樂巨頭達成協議以使用人工智慧。他希望確保「音樂版權持有者能夠因其對訓練資料的貢獻而獲得補償。UMG 和 YouTube 正在洽談授權藝術家的聲音和旋律來訓練 AI 模型。
亞馬遜 SageMaker	阿拉米	Alamy：公平傳播計劃的核心精神是致力於公平地報酬攝影師、藝術家和創作者。
	蓋蒂	Getty：我們遵守最高的道德標準，尊重他人的智慧財產權和隱私權。
	恩瓦托	
索尼		索尼董事長表示：「基於音樂創造的新產品和新業務必須在所有者的同意下進行開發，並給予適當的補償和認可。必須了解為什麼要進行人工智慧模型的訓練，訓練結果會發展出什麼產品，以及什麼樣的商業模式能夠讓藝術家的作品變得貨幣化。  「如果對生成式 AI 模型進行音樂訓練的目的是為了創作能夠在音樂市場上競爭的新音樂作品，那麼這種訓練就不屬於合理使用。在這種情況下，如果沒有藝術家和版權持有者的同意、認可和補償，培訓就無法進行。」

## 8. 專利

在編寫本報告時，我們確定了以下專利：

### 生成式人工智慧訓練作品報酬相關的專利

標題	參考	存款人所在國家	發表	先鋒
創意工作協作系統的系統與方法	US9100444	我們	磷	磷
合作專案中利害關係人股權評估	US10776889	我們	磷	磷
將生成的視覺內容歸因於訓練範例	US12033372	我們	磷	
基於模型的人工智慧生成內容歸因	US12013891	我們	磷	
生成作者向量	US10599770	我們	磷	
基於輸入的人工智慧生成內容歸因	US2024419949	我們		
人工智慧生成內容的產出歸因	US2024419720	我們		
費用分配系統及方法	US2023401550	我們		
內容的生成、使用與歸屬	WO2024166091	我們		
促進視覺內容的生成和使用	WO2023084512	我們		
基於聯邦學習的線上客戶選擇方法及系統、媒體及設備	CN118446287	中國	磷	
基於影像特徵相似性的影像索引生成	CN103207879	中國	磷	
影像局部特徵對整體特徵貢獻度評估方法及系統	CN111652239	中國	磷	
聯邦學習中量化客戶端貢獻的方法和系統	WO2023036184	中國		
基於Shapley值的聯邦學習行動裝置分佈資料處理方法	CN114912626	中國		
基於合作Sharpley值的聯邦學習激勵方法	CN115965092	中國		
具有局部預測器的聯邦學習方法和系統	WO2024031564	中國		
方法及協作系統	美國 9723059	歐盟	磷	磷
可解釋的自動編碼器和可解釋的生成對抗網路的方法	US11948083	歐盟	磷	
可解釋強化學習的架構	US11455576	歐盟	磷	
集體創作環境中的多層次貢獻可追溯性	US2015363743	歐盟		
協作創作獎勵方法及裝置	EP2958068	歐盟		

這份名單表明，在向權利人支付報酬方面，歐洲人並不比中國和美國落後，這與他們在人工智慧領域，尤其是生成式人工智慧領域的局外人地位形成了鮮明對比。

### The EU's position in complex (digital and green) technologies

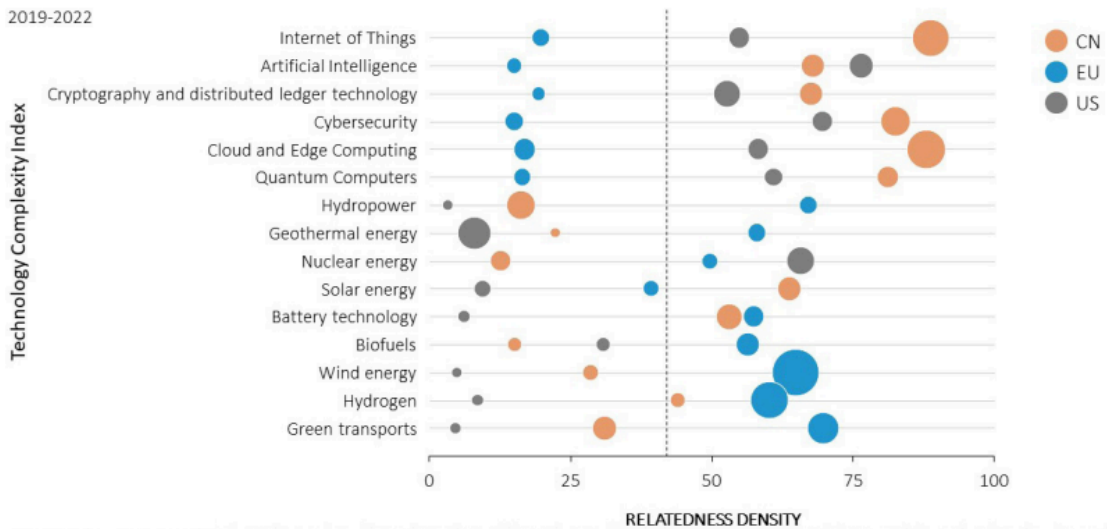


Table 7 Top patent owners in GenAI modes (companies), 2014–2023

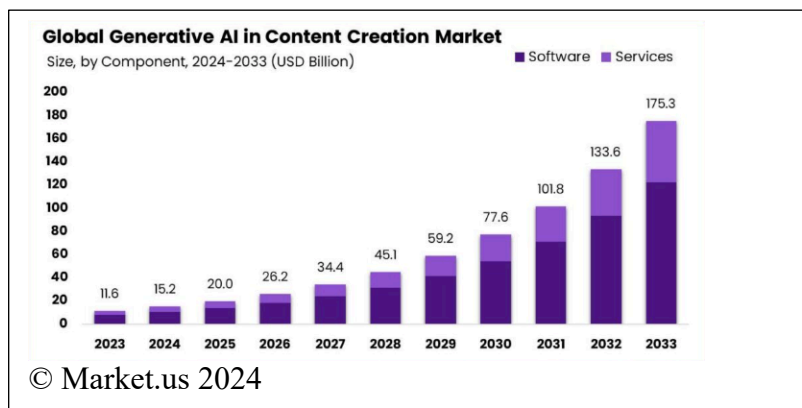
	Images/ video	Text	Speech/ voice/ music	3D image models	Molecules/ genes/ proteins	Software/code	Other modes
Tencent Holdings (China)	607	565	551	102	57	41	464
Ping An Insurance (China)	262	600	599	26	33	24	223
Baidu (China)	395	465	441	81	26	18	166
IBM (US)	101	274	168	13	37	42	132
Alibaba Group (China)	142	213	144	36	2	10	133
Samsung Electr. (Republic of Korea)	173	140	226	34	10	5	57
Alphabet/Google (US)	138	107	200	25	15	15	72
ByteDance (China)	173	82	112	21	6	7	90
Microsoft (US)	78	194	151	18	6	22	42
BBK Electronics (China)	125	58	117	12	0	4	116
Netease (China)	78	76	80	37	0	11	102
NTT (Japan)	37	100	169	11	0	3	62
Huawei (China)	104	84	72	24	1	4	102
China Mobile (China)	79	67	83	10	0	7	89
State Grid (China)	75	54	31	6	1	2	144
Adobe (US)	190	79	34	30	0	9	20
Sony Group (Japan)	87	35	105	42	0	1	42
Siemens (Germany)	124	25	8	10	3	4	59
Ant Group (China)	45	82	29	3	0	6	59
Industrial and Commercial Bank of China (China)	40	62	49	5	0	16	48

Source: WIPO, based on patent data from EconSight/IFI Claims, April 2024.



觀點

## 9 – 市場



僅就創造力領域而言，全球生成性人工智慧市場規模預計將在 2023 年達到 116 億美元，到 2031 年將超過 1,000 億美元。

舉個例子，15% 的公平報酬率將超過目前 OGC（集體管理組織）對藝術作品權利（音樂、視聽、現場表演、文學和視覺藝術）的徵收。

人工智慧訓練資料集的公司組織資料的收集、註釋、管理及其根據學習演算法的需求的適應。

全球人工智慧訓練資料集市場規模為 25 億美元，預計到 2032 年將達到 270 億美元。



### 生成式人工智慧資料作者的補償模型

商業模式	公司	作者報酬
刮痧 (1)	OpenAI、Midjourney、穩定性、MistralAI	不
數據的 Uber (2)	Mturk、ClickWorker、Gigwalk、Premise	訂單付款
內容庫(3)	Shutterstock、Getty、Canva、Adobe	付費培訓
媒體團體(4)	Axel Springer、金融時報、Stack Overflow、Reddit、世界報	不
公司 作者數量(5)	英國作家協會 由人類、人類原生人工智慧、經過公平訓練的資料集提供者聯盟創建	公平報酬

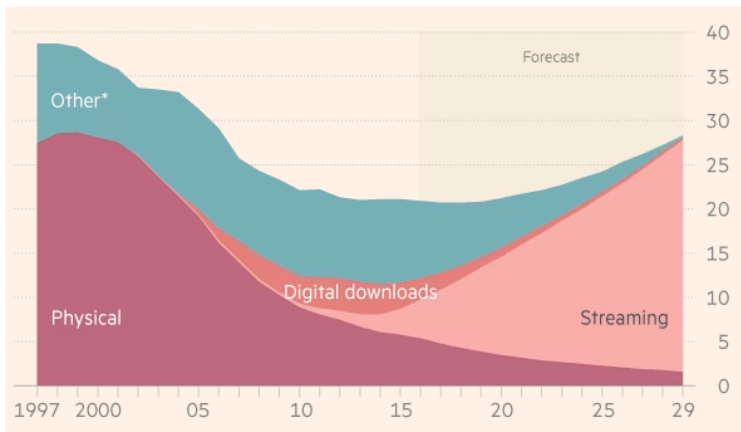
- 抓取：例如，DALL·E 首先使用來自互聯網的數億張帶字幕的圖像進行訓練。這種做法佔了大多數，並引發了無數訴訟。
- 數據的 Uber：人工智慧模型製定者已經開始對沖風險並確保數據供應鏈的安全，方式包括透過蓬勃發展的數據經紀行業以及為滿足需求而出現的與內容所有者達成的協議。

- 專門的人工智慧數據公司行業正在興起，它們創建了短期合約工網絡，以按需生產個人化內容。價格從每幀 1 至 2 美元不等，每部短片 2 至 4 美元不等，長片每小時 100 至 300 美元不等。文本的市場價格為每字 0.001 美元。照片、播客和醫療數據的所有者將獲得交易總金額的 20% 到 30% 左右的報酬。
- 內容庫：Shutterstock、Getty、Canva或 Adobe 擁有由作者提供的大型媒體庫，並根據出售的資料集獲得報酬。圖片領域的領導者Shutterstock已向 AI 供應商出售了超過 1 億美元的許可，佔其總收入的 12%，預計到 2024 年將增長 40 %。2030 年將達到100 億美元，在此期間每年成長率為22%。
- 媒體團體：資料庫擁有者授權資料集，無需向作者提供額外補償 NOTEREF \_Ref188784876 \h \\*  
MERGEFORMAT
- 作者協會：作者希望能夠定義其生成式人工智慧的授權選項。像作者協會這樣的想法正在蓬勃發展，透過人類平台創建，例如人類原生人工智慧，公平地經過培訓，資料集提供者聯盟或法國人工智慧委員會。現階段，經濟模型尚未定義，發展前景尚未量化。

### 生成式人工智慧作品授權提案

倡議	主要提案
作家協會	- 簽訂作品使用協議 - 支付使用受保護作品的版稅 - 數據使用透明化
人類創造	作品被貼上「人類創作」標籤 - 允許作者授權和控制許可證的平台
人類原生人工智慧	- 生成式人工智慧訓練資料集市場 - 追蹤每個使用內容的權限
訓練有素	- 標籤「相當 訓練」生成式人工智慧 - 人工智慧作品使用標準化合約
資料集提供者聯盟	- 公平的報酬 - 標準化合約(培訓付費、營業額特許權使用費、訂閱、行業許可) - 訓練資料使用的透明度 - 作者個人化授權
人工智慧委員會	- 公共平台 - 法國國家圖書館與國家視聽學院的中介

與音樂市場類似，我們目前正處於「Napster -IA」階段，等待與Spotify和Apple Music相當的解決方案。他們的商業模式從 2012 年開始淨化市場，而先前串流媒體和盜版一直被混淆。



© Financial Times, Goldman Sachs

同樣，我們認為，生成性人工智慧作品增值的潛在市場比 OGC 更大，這對權利人來說是一個機會。對稱地，生成式人工智慧操作員可以存取高品質的內容庫。

這種潛力目前使市場的其他部分受益，將透過「作者協會」類型的新經濟模式的提出和利益相關者的自願探索來釋放。

儘管這些舉措尚處於萌芽階段，但其不斷湧現為「流動財產」開闢了新的視角。

## 10 – 邁向液態財產

### Interview de Vincent Lorphelin

Le Soir (Bruxelles), Janvier 2024



「薩姆·奧特曼和埃隆·馬斯克表示，生成式人工智慧將使工作變得過時，因此要為全民收入做好準備。這與傑里米·里夫金幾年前提出的工作終結論不謀而合。這也與奇點理論相聯繫，奇點理論認為幾年之內，通用人工智慧去各方面都超越人類。證據就是我們一直說人工智慧永遠不會進入職業領域。富有創意（她永遠不會成為作家、設計師...）。只是從去年起我們就在那裡。堡壘的門被毀壞了。

相反，歐洲的DNA是社會市場經濟，是經濟創造社會。工作是有意義的。從歷史和戰略角度來看，歐洲的集體權利管理發達程度是美國的兩倍。因此，如果我們回到這個基本的問題上一般而言，為了平衡奇點論題，興趣一般的歐洲正在推廣模式分散化。

我們想像一下，這種去中心化的平台模式變得系統化，成為一個系統，就像今天Gafam這樣的中心化平台一樣。在這種情況下，操縱，這些都是智慧財產權，不是商品。這是非市場經濟價值。經濟學家卡爾·波蘭尼曾指出，市場經濟將勞動力商品化，但勞動力不能等同於商品。

中心化平台會盡可能賦予智慧財產權以市場價值。在去中心化的邏輯中，權利將以百分比來表示，而不是以貨幣來表示。經濟價值操縱的不再以歐元或美元表達。這種經濟不是以金錢為基礎，而是以財產為基礎。為了與埃米爾·左拉所說的「流動貨幣」（隨著財富不再來自土地所有權而是來自貨幣流通的經濟轉型）相比較，提出了「流動財產」的概念，其中營運資本經濟的一個關鍵點就是智慧財產權的流通。

今天，我們已經將一切都金融化，公司每天展示的帳目就像是在出售一樣，因為它用金錢來衡量一切。然而，有些無形的、表外的資產所表達的價值並不是用金錢來表達的。公司的財富遠超過金錢所能表達的範圍。當我們操縱液體特性時，我們會發現其他指標。我們正在逐漸走向一個讓另一種秩序的財富具有償付能力的社會。（點擊[此處了解更多](#)）。

正如Uber的平台模式啟發了“優步化”的概念一樣，我們將版權社會經濟模式的傳播現象稱為“版權化”，這種模式將透過流動財產來增強人力資本。

## 11. 提升人類智力

總裁布拉德史密斯正在對公司犯下道德錯誤。在談到ChatGPT等生成式人工智慧時，他說：「根據版權法，我們都有閱讀和學習的權利。我們現在問自己是否可以讓機器以同樣的方式學習。方式。我認為有必要 社會才能使這一切成為可能。」

這次幹預也許看起來微不足道。畢竟，如果機器能夠以極低的價格吸收網路來創作文章和插圖，我



們為什麼要剝奪自己呢？《紐約時報》和其他反對如此使用其內容的版權所有者難道沒有為微軟帶來不利的審判嗎？如果機器不抄襲，會有什麼危害？這是這場演講令人無法接受的結果。

### 有利的法律模糊性

Gafam平台收穫了大量的小額捐款，並且促進了免費勞動。他們再次透過人工智慧改變維度，人工智慧的貪婪要求他們征服新的領土，也就是作者的領土。

作為例外，立法者定義了“合理使用”，授權谷歌發布私人內容的摘錄。由於法庭的結果仍不確定，前律師布拉德史密斯知道公眾輿論將發揮重要作用。這就是為什麼他要提煉「機器學習的權利」這個摻假的心理形象。他的「命令「社會」只是作者捍衛自己領土的合法性。

他透過辯論虛假的道德問題來轉移注意力。人類智力是否注定降級？我們可以讓人工智慧做出決策嗎？他很可能在埃隆面前擺出一個智者的姿態 馬斯克不那麼客氣，他們都是同一個幻想的預言者：超越人類智慧。

### 優越感的假象？

最好的小說永遠無法取代人生經驗。最好的「視訊通話」也永遠無法取代真正的會議。無論模型多麼複雜，它都不是實體。我們每個人都擁有一種對世界的智能，這種智能將永遠超越機器，而機器除了其模型之外沒有其他智能。

人工智慧啟迪我們的智慧。它們使環保警察能夠發現污染源、使老師能夠了解一千名學生的困難、或使實驗室能夠預測某種藥物的效果。



傳播機器即將超越的幻想是不負責任的。這是對嚴格意義上的人類勞動的否定，人類勞動的靈感恰來自於世界及其模型，當藝術家揭示一些未說出口的東西時，當駕駛員彌補反應不佳的槓桿時，當醫生懷疑一個過於明顯的診斷時，或當法官有著堅定的信念。

### 作者辯護

競爭的幻想，並將人權歸於機器，這是對社會的違背。相反，我們必須肯定人類智慧的卓越地位，並保證其領土的完整性。

沒有這位世界頂級人工智慧人物，布拉德史密斯的幻想就毫無意義。受這種幻想所迷惑，人工智慧部際委員會斷言機器將超越人類。幸運的是，她捍衛了人類創造的公平報酬，但沒有意識到這是對人類創造的欣賞 即使這種準確性是被操縱的。

歐洲已經決定透過AI ACT來保護版權。現在它必須建立AI所開發領土的土地登記冊並維護其租金。  
(詳情[請閱讀](#)2024年5月6日《迴聲報》)



## 12 –歐洲的薩米化解決方案

### Entretien avec Jean-Paul Betbèze

Economiste, Membre du Comité scientifique de la  
Fondation Robert Schuman



### 「清真化」的重點應該是什麼？

薩米化』的解決方案應該：

- 像 OGC 一樣，基於作者的知識產權。有必要維護和保障作者的智慧財產權，否則在當前的通訊革命中，智慧財產權可能會被消除或消失。通訊正在變得全球化，它是我們正在經歷的重大變化的基礎，資訊在各個方向上成倍增加。每個人的貢獻都必須得到承認，除非我們為了某些作者的利益而過度集中精力，從而扼殺當前的全球運動，而其他作者在其工作和思想中得不到任何優勢，得不到認可，得不到報酬。由於資本可用性方面的差距，這在很大程度上對美國有利，危險在於表面上的創造來源被壟斷。

- 收取人工智慧作品開發使用的使用費。「Sacemization」不僅使我們能夠追

踪，而且還能識別和估計來源，即使它們看起來很微不足道。這個過程使得所有的起源得以保留，從而使得所有創造和改進得以認可：沒有失去任何東西，一切都得以保留。

- 識別創造力並將其與業餘內容區分開來。在作品的擴散中，能凸顯出原創作品、加以宣傳、加以推廣是十分重要的。
- 實行個別激勵機制，使創意價格在市場上具有競爭力，並與生產成本相容。作品增多的風險是微不足道的，尤其是考慮到它們的實體生產成本明顯較低。僅與生產成本掛鉤的計費可能會降低動力。「薩米化」的巨大優勢在於擺脫與生產成本掛鉤的價格邏輯，轉向另一種邏輯，即市場邏輯，這種邏輯基於成功的重要性和資訊的重複。

### 「清真化」驅動的市場有什麼優勢？

依照這個市場邏輯，衡量需求，進而產生預期需求，進而創造新價格，進而形成利潤。在這種背景下，利潤率只能增加，因此邊際利潤也會增加，估值也會增加。搜尋引擎偵測到的低成本與透過 Sacem 類工具更新的具有客觀基礎的估值之間的差異是所提出邏輯的基礎。它將使每個人的貢獻都得到凸顯，特別是在歐洲背景下。

## 對於創造性活動呢？

我們不要忘記，壟斷競爭始終在起作用：它推動競爭對手的數量減少，直到只出現一個：這就是壟斷。在那裡，比起其他任何地方，勝利者都能贏得所有的賭注。「薩西姆化」必須透過在單一壟斷者出現之前實現更好的報酬分配來反對這種做法。這顯然有利於研究、創新和創造，因為我們知道壟斷局面對壟斷者本身不利，從長遠來看，市場局面會枯竭，因為其動態會導致其利潤下降。因此，「薩西姆化」尊重創造，並減緩壟斷本身無情的、最終是破壞性的選擇。

在思想市場中，速度是致命的，對貢獻的記憶具有決定性作用，同時也與其公平的報酬結合。今天這已經成為可能，而且可以理解的是，比以往任何時候都更有必要。它是關於透過技術進步來拯救創造。»

Michelle Bergadaà, Professeure émérite à l'université de Genève.

Co-auteurice de *Réinventer l'intégrité académique à l'ère de l'intelligence artificielle* (à paraître, EMS, février 2025)



你說《人工智慧法案》不僅涉及作者，還涉及研究人員，為什麼？

MB：“我們需要考慮適用《人工智慧法案》來補償‘真實研究’的作者。”法國給研究人員的薪水很低，這會減緩他們出國或進入私部門的速度。用官僚主義壓制年輕研究人員比給他們優厚報酬對社會造成的成本大得多。許多人變得沒有動力，花額外的時間去教學而不是做研究。因此，如果國家不能支付他們適當的報酬，那麼出版文章、書籍和報告的體系就必須這樣做。這必須成為歐洲生成式人工智慧策略的基石。還附帶自動詐欺偵測系統的額外好處。

詐欺不是一直都存在嗎？

美國的鴉片類藥物危機 [2021 年，光是芬太尼及相關處方藥就造成了約 71,000 人因藥物過量死亡] 源於一篇虛假、被錯誤引用和轉載的文章。這遠在人工智慧之前就已存在。在人工智慧的幫助下，欺詐性文章的傳播主要透過新文章進行。這是一種新的群假模式。

如今三種時間並存。首先，（川普和 Z 世代的）瞬間和情緒化時間不僅允許虛假訊息的產生，還允許虛假作者的產生，而虛假作者同樣可信。其次，昔日社群的永恆現在維持著一個另類世界，

在這個世界裡，地球是平的，漂白水可以消滅新冠病毒。第三線性時間由伊隆的物質技術驅動 麝香。人工智慧是一種詐欺加速器，它在時間的間隙中蓬勃發展。

怎樣行動？

Pubpeer和Retraction Watch Database致力於逐一抓獲違法者。但完成調查並在報紙上撤回報道花了三年時間。

必須有系統地解決這一現象，使用誠信雷達、強制詐欺者退還所收到的錢以及排除人工智慧資料庫。我們社會的未來在於我們所教育的年輕人。讓我們採取堅定的立場並透過生成式人工智慧獎勵他們的原創作品來幫助他們。這將激勵許多年輕、誠實的研究人員創造知識。 »

建議

## 13 – Controv3rse 解決方案

Open AI 創辦人 Sam Altman 希望擁有「能讓創作者獲得收入的新商業模式。」»

我們看到，資料模型的Uber的出現，避免了生成式人工智慧的退化，但卻對人力資本構成了威脅。經濟模式必須依靠OGC，透過流動性和資本化來開啟這種資本重估的視角。

我們向 Sam Altman 和其他生成式 AI 業者推薦的商業模式是這樣的：

### Controv3rse 解決方案

推薦	開放人工智慧	奧格斯堡
1 - 保留	保留 15% 的生成性人工智慧營業額，用於公平支付作者和版權持有者的報酬	為作者提供一個平台，在適當的情況下，反對生成式人工智慧使用其作品，或逐一作品選擇許可證
2 - 細分	按市場細分確定主要的 OGC（媒體/資料集*國家）	佔領一個細分市場（媒體/資料集*國家）
3 - 方法	透過 OGC 發布歸因方法，詳細說明恢復時的消歧方法	就 OGC 之間的協調或互惠協議進行談判
4 - 不歸因	按國家/地區識別未分配金額的基金	基金理賠管理
5 - 報告	發布足夠詳細的摘要  向每個已識別的 OGC 發送一份月度報告，其中包含作者的識別碼、其作品使用的詳細資訊、合成內容以及歸因的價值	向操作員發送無法被生成式 AI 使用的作品列表
6 - 付款	將儲備金分配給 OGC 和基金	將所得分配給作者和版權持有者





## 公平報酬

透過生成式人工智慧實現公平補償的做法可以受到Spré案例的啟發。在法國，該組織根據夜總會的營業額收取版稅，用於造福音樂家。其中 120 個代表性樣本配備了由Sacem和Spré資助的「Yacast」盒子。Yacast收集串流音樂的播放列表，並按藝術家、唱片公司、流派和播出週數發布統計數據。在法定扣除 9% 後，Spré將稅款分配給 OGC（SCPA 50%、ADAMI 25% 和 SPEDIDAM 25%），並將Yacast統計數據傳達給他們，以補償最終受益者（表演藝術家、製作人、一般利益行動）。

## 公平利率

在其2020年的報告《作者與創作行為》中對「作者的報酬比例只有在特殊情況下才能達到或超過營業收入的10%」這一事實表示遺憾。他將這一觀察歸因於「藝術家-作者與下游參與者（出版商、分銷商、製作人）之間的關係似乎極不平衡，這導致他在許多情況下對創作理念本身產生了質疑。」契約自由。為了解決這個問題，他希望“通過迅速組織專業選舉來加強藝術家和作家的集體力量，這將為即將成立的國家藝術家和作家委員會提供實質和合法性，以作為與廣播公司進行集體談判”，建立「一種由贊助商優先支付作者工作時間報酬的合約制度」。

由於重新評估勞動價值的想法並不適用於生成式人工智慧的情況，我們建議將公平費率定義為所有作者貢獻對生成內容商業形象的效用價值。這就是 Shapley 值所依據的原理，經濟學家恩斯特·費爾 (Ernst Fehr)估計，新聞為 Google 搜尋帶來的價值為 14%。自從有了一個著名的問題“這個評論對你有用嗎？”» 由亞馬遜推出，所有商業網站都透過增加現在普遍使用的實用指標來提高其產品的品質。

將稅基設定為所產生內容的營業額，將使作者的公平報酬不依賴於該行業的或多或少的垂直細分，例如：作者/平台/資料集/通用模型/微調/介面/一代。

## 15 – 許可 – 致謝

本說明及其表格可在 CC-BY 許可下訪問，但須遵守由協會 Controv3rse 授權的生成性人工智慧對其開發成果的公平分享，編號為 RNA W751267044。

本文檔由 Google 翻譯自動翻譯。如有歧義，以法文版本為準。

法案》部長級代表團成員、獨立智庫 Controv3rse 和 Institut de l' Economie 的成員以及其他因其專業知識而受到認可的人士之間的反思和討論的靈感。它擴展了先前為Pradal -Rimbaud議會代表團進行的研究以及在《世界報》和《迴聲報》上發表的專欄文章。這些智庫的創始人兼聯合總裁、本文作者文森特·洛菲林 (Vincent Lorphelin) 熱烈感謝他們的想法和討論的高品質。



議會代表團和部長代表團團長和報告員：

Alexandra Bensamoun) ， 巴黎薩克雷大學法學教授， 前人工智慧委員會成員

Joëlle Farchy ， 大學教授， m2ecn 主任和PcEn主席  
巴黎第一大學， CSPLA 成員

( Philippe Pradal) ， 國民議會濱海阿爾卑斯省第三選區議員

Stéphane Rambaud, 國民議會瓦爾省第三選區議員

國務委員會請求主管

Julie Groffe -Charrier, 讓莫內學院私法 HDR 講師

布萊恩 (Bastien Blain), 巴黎第一大學經濟學副教授、 數據與行為科學家

各位專家、各位企業家：

威廉貝利

Bolero Music共同創辦人兼總裁

埃馬紐埃爾·貝納澤拉

Jolibrain董事長兼首席執行官

讓-保羅·貝貝茲

Betbèze Conseil經濟學家

米歇爾·貝爾加達

學術詐欺與剽竊研究與行動研究所所長

米歇爾·博科布扎

巴黎學院任務經理

洛朗·切爾沃尼

Talan研究與創新總監

查爾斯·埃里克·德拉夏佩勒

Myriad Data創辦人

尚菲利浦·克萊爾

Keyrus行銷、溝通與創新總監

艾莉絲·科塔勒姆

CogNeed共同創辦人、巴黎第九大學

PSL 副教授

克里斯托夫·科萊特

AskLocala創辦人

弗雷德里克·達揚

Exactcure創辦人

紀堯姆·德沃

Aleia共同創辦人兼 AI Cargo Foundation  
管理員

喬納森·多利

Live Crew創辦人

威廉·艾爾丁

創辦人， XXII

史蒂芬妮·弗萊徹

Logion Network共同創辦人

拉斐爾·弗里施

HawAI.tech聯合創辦人

阿蘭·加尼埃

Jamespot首席執行官

安東尼·格雷夫林

Disaitek創始人

菲利普·古根  
總統，索爾加-地圖 乳液

法蘭西斯·哈赫姆  
Codenekt創辦人

羅道夫·哈塞爾萬德  
蛙創辦人 機器人

布萊斯·霍勞  
Semdee創辦人

馬蒂亞斯·霍利爾  
Wintics共同創辦人

凱西·喬利  
律師， IPso

讓·拉特熱  
Oktal -SE創辦人

弗雷德里克·勒費弗爾-納雷  
Niji人工智慧資料總監

尤尼斯 萊姆拉貝特  
Everysens創辦人

雅克·萊維-維埃勒  
分析總裁、法國國家資訊與自動化研究所研究主任

奧利維爾·拉博德  
BPCE 創新與數位轉型負責人

勒弗雷斯蒂爾  
Panodyssey創辦人

西斯汀·洛菲林  
AI 工程師， UTC， INSEAD 學生

艾梅里克 馬蘇雷勒  
Spoon.ai聯合創辦人

皮埃爾·米拉萊斯  
Footovision共同創辦人

艾曼紐爾·莫伊朗  
France Meta 聯合創辦人

克萊門特·梅維爾  
Manzalab創辦人- Teemew

愛德華·米奧利斯  
Polycube創辦人

娜塔莉·內維簡斯  
阿爾多瓦大學私法與人工智慧教授

雷米·奧茲坎  
Crypto4All 共同創辦人

-雅克·基斯夸特  
魯汶理工學院密碼學教授、《世界報》專欄作家

讓-米歇爾·薩洛蒙  
遊戲作者協會會長

弗雷德里克·蘇夫萊  
Haapie共同創辦人

克萊門特·特基  
Ternoia共同創辦人

弗朗索瓦-澤維爾·托朗斯  
Vaultys執行長

阿諾·圖阿蒂  
助理律師， #Hashtag

克里斯多福·特里科  
La Forge 共同創辦人

基里安·韋爾默施  
Golem.ai 聯合創始人兼首席執行官

# Bibliographie

<sup>1</sup> <https://www.hollywoodreporter.com/business/business-news/openai-ceo-sam-altman-content-creators-elon-musk-1236077011/>

<sup>2</sup> <https://www.hollywoodreporter.com/business/business-news/openai-ceo-sam-altman-content-creators-elon-musk-1236077011/>

<sup>3</sup> [https://www.lemonde.fr/economie/article/2024/02/26/mistral-ai-en-ia-nous-pouvons-rivaliser-avec-google-ou-openai\\_6218645\\_3234.html](https://www.lemonde.fr/economie/article/2024/02/26/mistral-ai-en-ia-nous-pouvons-rivaliser-avec-google-ou-openai_6218645_3234.html)

<sup>4</sup> <https://petapixel.com/2022/12/21/midjourney-founder-admits-to-using-a-hundred-million-images-without-consent/>

<sup>5</sup> <https://a16zcrypto.com/posts/article/read-write-own-intro/>

<sup>6</sup> <https://www.buildingtheopenmetaverse.org/episodes/creator-economy>

<sup>7</sup> <https://beincrypto.com/ai-companies-solve-conflict-copyright-claims/>

<sup>8</sup> <https://www.culture3.com/posts/why-the-internet-needs-a-metaverse-magna-carta-moment-with-animoca-brands-yat-siu-digital-property-rights>

<sup>9</sup> voir à 25:30

<sup>10</sup> <https://techcrunch.com/2024/06/25/backed-by-david-sacks-garry-tan-and-walter-isacson-created-by-humans-helps-authors-license-their-creative-work-to-ai-models/> , <https://www.createdbyhumans.ai/manifesto>

<sup>11</sup> <https://x.com/ednewtonrex/status/1770850313492221986> ; l'utilisation de documents protégés par le droit d'auteur de manière transformatrice n'équivaut pas à une violation du droit d'auteur <https://media.shellypalmer.com/wp-content/images/2023/08/Tremblay-v-OpenAI-Motion-to-Dismiss-8-30-2023.pdf>

<sup>12</sup> Extraits de l'AI act

(97)

La notion de modèles d'IA à usage général devrait être clairement définie et distincte de la notion de systèmes d'IA afin de garantir la sécurité juridique. La définition devrait se fonder sur les principales caractéristiques fonctionnelles d'un modèle d'IA à usage général, en particulier la généralité et la capacité d'exécuter de manière compétente un large éventail de tâches distinctes. Ces modèles sont généralement entraînés avec de grandes quantités de données, au moyen de diverses méthodes, telles que l'apprentissage auto-supervisé, non supervisé ou par renforcement. Les modèles d'IA à usage général peuvent être mis sur le marché de différentes manières, notamment au moyen de bibliothèques, d'interfaces de programmation d'applications (API), de téléchargements directs ou de copies physiques. Ces modèles peuvent être modifiés ou affinés et ainsi se transformer en nouveaux modèles. Bien que les modèles d'IA soient des composants essentiels des systèmes d'IA, ils ne constituent pas en soi des systèmes d'IA. Les modèles d'IA nécessitent l'ajout d'autres composants, tels qu'une interface utilisateur, pour devenir des systèmes d'IA.

(107)

Afin d'accroître la transparence concernant les données utilisées dans le cadre de l'entraînement préalable et de l'entraînement des modèles d'IA à usage général, y compris le texte et les données protégés par la législation sur le droit d'auteur, il convient que les fournisseurs de ces modèles élaborent et mettent à la disposition du public un résumé suffisamment détaillé du contenu utilisé pour entraîner les modèles d'IA à usage général. Tout en tenant dûment compte de la nécessité de protéger les secrets d'affaires et les informations commerciales confidentielles, ce résumé devrait être généralement complet en termes de contenu plutôt que détaillé sur le plan technique afin d'aider les parties ayant des intérêts légitimes, y compris les titulaires de droits d'auteur, à exercer et à faire respecter les droits que leur confère la législation de l'Union, par exemple en énumérant les principaux jeux ou collections de données utilisés pour entraîner le modèle, tels que les archives de données ou bases de données publiques ou privées de grande ampleur, et en fournissant un texte explicatif sur les autres sources de données utilisées. Il convient que le Bureau de l'IA fournisse un modèle de résumé, qui devrait être simple et utile et permettre au fournisseur de fournir le résumé requis sous forme descriptive.

(108)

En ce qui concerne l'obligation imposée aux fournisseurs de modèles d'IA à usage général de mettre en place une politique visant à respecter la législation de l'Union sur le droit d'auteur et de mettre à la disposition du public un résumé du contenu

utilisé pour l'entraînement, le Bureau de l'IA devrait vérifier si le fournisseur a rempli cette obligation sans vérifier ou évaluer œuvre par œuvre les données d'entraînement en ce qui concerne le respect du droit d'auteur. Le présent règlement n'affecte pas l'application des règles en matière de droit d'auteur prévues par la législation de l'Union.

#### Article 53

Obligations incombant aux fournisseurs de modèles d'IA à usage général

1. Les fournisseurs de modèles d'IA à usage général:

d) élaborent et mettent à la disposition du public un résumé suffisamment détaillé du contenu utilisé pour entraîner le modèle d'IA à usage général, conformément à un modèle fourni par le Bureau de l'IA.

#### Article 56

Codes de bonne pratique

2. Le Bureau de l'IA et le Comité IA s'efforcent de veiller à ce que les codes de bonne pratique couvrent au moins les obligations prévues aux articles 53 et 55, y compris les questions suivantes:

b) le niveau approprié de détail pour le résumé du contenu utilisé pour l'entraînement;

[https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=OJ:L\\_202401689](https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=OJ:L_202401689)

[https://www.lemonde.fr/economie/article/2024/09/14/les-contenus-utilises-pour-entraîner-les-ia-ont-une-valeur-et-un-prix\\_6317395\\_3234.html](https://www.lemonde.fr/economie/article/2024/09/14/les-contenus-utilises-pour-entraîner-les-ia-ont-une-valeur-et-un-prix_6317395_3234.html)

<sup>13</sup> <https://x.com/ednewtonrex/status/1848741876662436078>

<sup>14</sup> [https://www.culture.gouv.fr/Media/medias-creation-rapide-ne-pas-supprimer/note\\_etape\\_economique\\_mission\\_cspla\\_remuneration\\_ia\\_dec\\_2024.pdf](https://www.culture.gouv.fr/Media/medias-creation-rapide-ne-pas-supprimer/note_etape_economique_mission_cspla_remuneration_ia_dec_2024.pdf)

<sup>15</sup> <https://www.clickworker.com/>

<sup>16</sup> <https://www.businessinsider.com/gigwalk-iphone-app-2011-5> , <https://premise.com/blog/the-new-and-improved-premise-app-launches/>

<sup>17</sup> <https://www.mckinsey.com/mgi/our-research/generative-ai-and-the-future-of-work-in-america>

<sup>18</sup> <https://www.lesechos.fr/tech-medias/hightech/video-levee-de-fonds-record-pour-cette-start-up-de-lia-2138620>

<sup>19</sup> <https://www.cae-eco.fr/staticfiles/pdf/097.pdf> p.117

<sup>20</sup> <https://www.vincentlorphelin.com/blog/les-echos-ia-faut-il-attribuer-des-droits-humains-aux-machines?categoryId=33952> , <https://www.theverge.com/2024/9/25/24254042/mark-zuckerberg-creators-value-ai-meta>

<sup>21</sup> [https://www.oecd-ilibrary.org/education/l-investissement-dans-le-capital-humain\\_9789264262898-fr](https://www.oecd-ilibrary.org/education/l-investissement-dans-le-capital-humain_9789264262898-fr)

Miroir de l'humanité et de l'esthétique, fondement de nos identités et de nos communautés, c'est une source infinie d'inspiration et d'innovation pour l'économie et le développement durable [https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32022G1207\(01\)](https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32022G1207(01))

<sup>22</sup> L'Union européenne voulait « devenir l'économie de la connaissance la plus compétitive et la plus dynamique du monde, capable d'une croissance économique durable accompagnée d'une amélioration quantitative et qualitative de l'emploi et d'une plus grande cohésion sociale ».

<sup>23</sup> Jean Tirole. Economie du Bien Commun, p. 566. PUF, 2016

<sup>24</sup> Jean Tirole. Economie du Bien Commun, pp. 569 et 583. PUF, 2016

<sup>25</sup> Jean Tirole. Economie du Bien Commun, p. 567. PUF, 2016

<sup>26</sup> Jean Tirole. Economie du Bien Commun, p. 596. PUF, 2016

<sup>27</sup> Jean Tirole. Economie du Bien Commun, p. 574. PUF, 2016

<sup>28</sup> Jean Tirole. Economie du Bien Commun, p. 576. PUF, 2016

<sup>29</sup> <https://arxiv.org/pdf/1904.02868> , <https://arxiv.org/pdf/1902.10275>

<sup>30</sup> [https://www.researchgate.net/publication/334964730\\_A\\_New\\_Deep\\_Learning\\_Based\\_Multi-Spectral\\_Image\\_Fusion\\_Method](https://www.researchgate.net/publication/334964730_A_New_Deep_Learning_Based_Multi-Spectral_Image_Fusion_Method)

<sup>31</sup> <https://arxiv.org/html/2406.14862v1>

<sup>32</sup> [https://arxiv.org/abs/1811.00982?utm\\_source=chatgpt.com](https://arxiv.org/abs/1811.00982?utm_source=chatgpt.com)

<sup>33</sup> 136 millions d'images sous licence CC BY en 2010 <https://creativecommons.org/2010/03/10/creative-commons-licenses-on-flickr-many-more-images-slightly-more-freedom/> contre 4 milliards d'images au total <https://fr.wikipedia.org/wiki/Flickr>

- 34 <https://www.flickrhelp.com/hc/en-us/articles/4404078674324-Change-Your-Photo-s-License-in-Flickr>
- 35 [https://shs.cairn.info/revue-reseaux-2009-2-page-91?lang=en&utm\\_source=chatgpt.com](https://shs.cairn.info/revue-reseaux-2009-2-page-91?lang=en&utm_source=chatgpt.com)
- 36 [https://www.researchgate.net/publication/255726672\\_The\\_'Re-Identification'\\_of\\_Governor\\_William\\_Weld's\\_Medical\\_Information\\_A\\_Critical\\_Re-Examination\\_of\\_Health\\_Data\\_Identification\\_Risks\\_and\\_Privacy\\_Protections\\_Then\\_and\\_Now](https://www.researchgate.net/publication/255726672_The_'Re-Identification'_of_Governor_William_Weld's_Medical_Information_A_Critical_Re-Examination_of_Health_Data_Identification_Risks_and_Privacy_Protections_Then_and_Now)
- 37 [https://www.cs.utexas.edu/~shmat/shmat\\_oak08netflix.pdf](https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf) , <https://arxiv.org/html/2312.15489v2> , <https://www.sciencedirect.com/science/article/pii/S1877050918303260>
- 38 [https://pubmed.ncbi.nlm.nih.gov/?term=Johnson+JA&cauthor\\_id=28805569](https://pubmed.ncbi.nlm.nih.gov/?term=Johnson+JA&cauthor_id=28805569)
- 39 <https://huggingface.co/datasets/bookcorpus/bookcorpus>
- 40 [https://huggingface.co/datasets/stas/openwebtext-10k/viewer/plain\\_text/train?q=written+by&row=4593](https://huggingface.co/datasets/stas/openwebtext-10k/viewer/plain_text/train?q=written+by&row=4593)
- 41 <https://huggingface.co/datasets/c17hawke/stackoverflow-dataset>
- 42 <https://stackoverflow.com/users/207072/sc-ray>
- 43 <https://stackexchange.com/users/71587/sc-ray>
- 44 <https://commonvoice.mozilla.org/en/datasets>
- 45 <https://analyzer.cv-toolbox.web.tr/examine/en/18.0>
- 46 <https://tollbit.com/bot-paywall/>
- 47 <https://www.businesswire.com/news/home/20241209071998/en/ProRata-Announces-Gist.ai-New-AI-Search-Engine-Based-Entirely-on-High-Quality-Licensed-Content>
- 48 [https://musicworks.substack.com/p/bringing-responsible-and-explainable?utm\\_campaign=post&utm\\_medium=web](https://musicworks.substack.com/p/bringing-responsible-and-explainable?utm_campaign=post&utm_medium=web) ; brevets <https://patents.google.com/patent/US12013891B2> <https://www.freepatentsonline.com/20240419949.pdf>
- 49 <https://www.adobe.com/fr/products/firefly.html>
- 50 <https://www.gist.ai/>
- 51 <https://www.cnbc.com/video/2024/08/07/ai-startup-inks-content-media-deal.html> , [https://www.youtube.com/watch?v=7F\\_ZsN-h5mg](https://www.youtube.com/watch?v=7F_ZsN-h5mg) , <https://www.wired.com/story/bill-gross-prorata-generative-ai-business/>
- 52 <https://blog.bria.ai/blogs/bria-raises-24m-series-a-to-pioneer-responsible-visual-generative-ai-for-enterprises>
- 53 <https://tollbit.com/blog/series-a/>
- 54 <https://arxiv.org/pdf/2406.11857>
- 55 <https://www.alltageinesfotoproduzenten.de/2023/05/08/analyse-wie-viel-zahlt-shutterstock-fuer-ki-trainingsdaten-an-anbieter/>
- 56 <https://www.theverge.com/2022/10/25/23422359/shutterstock-ai-generated-art-openai-dall-e-partnership-contributors-fund-reimbursement>
- 57 <https://www.lexpress.fr/economie/high-tech/quand-la-boite-noire-des-ia-generatives-livre-ses-secrets-BFUANKGCOZF2DDGJRNFDNAVHZU/>
- 58 <https://www.nvidia.com/en-us/on-demand/session/siggraph2024-sigg2412/>
- 59 <https://blogs.microsoft.com/blog/2024/10/01/an-ai-companion-for-everyone/> , [https://aboutus.ft.com/press\\_release/openai](https://aboutus.ft.com/press_release/openai) , <https://www.axelspringer.com/en/ax-press-release/axel-springer-and-openai-partner-to-deepen-beneficial-use-of-ai-in-journalism>
- 60 <https://www.wired.com/story/bill-gross-prorata-generative-ai-business/>
- 61 [https://worldwide.espacenet.com/searchResults?submitted=true&locale=en\\_EP&DB=EPODOC&ST=advanced&TI=&AB=&PN=&AP=&PR=&PD=&PA=idealab&IN=&CPC=&IC=](https://worldwide.espacenet.com/searchResults?submitted=true&locale=en_EP&DB=EPODOC&ST=advanced&TI=&AB=&PN=&AP=&PR=&PD=&PA=idealab&IN=&CPC=&IC=)
- 62 [https://worldwide.espacenet.com/publicationDetails/biblio?DB=EPODOC&II=0&ND=3&adjacent=true&locale=en\\_EP&FT=D&date=20241219&CC=US&NR=2024419949A1&KC=A1](https://worldwide.espacenet.com/publicationDetails/biblio?DB=EPODOC&II=0&ND=3&adjacent=true&locale=en_EP&FT=D&date=20241219&CC=US&NR=2024419949A1&KC=A1)
- 63 <https://arxiv.org/pdf/2404.13964v2> , <https://arxiv.org/abs/2404.13964>
- 64 <https://arxiv.org/pdf/2406.11011>
- 65 <https://www.technologyreview.com/2024/03/04/1089403/large-language-models-amazing-but-nobody-knows-why/> , <https://openai.com/openai-o1-contributions/>

<sup>66</sup> The Pythia suite is a suite of LLMs that satisfies three key properties: 1. Models span several orders of magnitude of model scale. 2. All models were trained on the same data “The Pile” in the same order. 3. The data and intermediate checkpoints are publicly available for study. <https://arxiv.org/pdf/2304.01373>

The Pile is a 825.18 GiB English text dataset designed for training large scale language models. The Pile is composed of 22 diverse and high-quality datasets, including both established natural language processing datasets and several newly introduced ones. <https://arxiv.org/pdf/2101.00027>

<sup>67</sup> <https://arxiv.org/pdf/2404.13964>

<sup>68</sup> <https://arxiv.org/pdf/2303.14186>

<sup>69</sup> <https://arxiv.org/abs/2306.09345>

<sup>70</sup>

[https://openaccess.thecvf.com/content/CVPR2024/papers/Asnani\\_ProMark\\_Proactive\\_Diffusion\\_Watermarking\\_for\\_Causal\\_Attribution\\_CVPR\\_2024\\_paper.pdf](https://openaccess.thecvf.com/content/CVPR2024/papers/Asnani_ProMark_Proactive_Diffusion_Watermarking_for_Causal_Attribution_CVPR_2024_paper.pdf)

<sup>71</sup> <https://arxiv.org/html/2406.09408v1>

<sup>72</sup> <https://arxiv.org/pdf/2109.09046>

<sup>73</sup> [https://www.researchgate.net/publication/334964730\\_A\\_New\\_Deep\\_Learning\\_Based\\_Multi-Spectral\\_Image\\_Fusion\\_Method](https://www.researchgate.net/publication/334964730_A_New_Deep_Learning_Based_Multi-Spectral_Image_Fusion_Method)

<sup>74</sup> <https://arxiv.org/pdf/2404.13964>

<sup>75</sup> <https://arxiv.org/pdf/2406.11011>

<sup>76</sup> <https://arxiv.org/pdf/2312.06646>

<sup>77</sup> Monte Carlo, In-Run Data Shapley <https://arxiv.org/pdf/2406.11011>

DU-Shapley <https://arxiv.org/pdf/2306.02071>

Group testing, All-S influence, Largest-S influence <https://arxiv.org/pdf/1902.10275>

<sup>78</sup> <https://arxiv.org/pdf/1902.10275>

<sup>79</sup> Getty-backed AI image generator BRIA snags fresh funding. Katie Paul, Reuters, February 21, 2024.

<https://www.reuters.com/technology/getty-backed-ai-image-generator-bria-snags-fresh-funding-2024-02-21/>

<sup>80</sup> <https://investor.shutterstock.com/news-releases/news-release-details/shutterstock-launches-first-ethical-generative-3d-api>

<sup>81</sup> June 20, 2023 [Nvidia](https://www.nvidia.com/en-us/generative-ai/)

<sup>82</sup> <https://community.adobe.com/t5/stock-contributors-discussions/announcing-bonus-payment-for-adobe-firefly-training/td-p/14081556>

<sup>83</sup> <https://www.adobe.com/products/firefly.html>

<sup>84</sup> <https://newsroom.gettyimages.com/en/istock/getty-images-launches-generative-ai-by-istock-for-small-businesses-designers-and-marketers>

<sup>85</sup> <https://www.prnewswire.com/news-releases/shutterstock-expands-long-standing-relationship-with-meta-301719769.html>

<sup>86</sup> <https://www.shutterstock.com/blog/shutterstock-ai-image-generator-dall-e-3>

<sup>87</sup> <https://www.prnewswire.com/news-releases/shutterstock-joins-forces-with-lg-ai-research-to-advance-ai-technology-to-revolutionize-the-creative-journey-301680992.html>

<sup>88</sup> [https://www.judiciary.senate.gov/imo/media/doc/2023-05-16\\_-\\_qfr\\_responses\\_-\\_altman.pdf](https://www.judiciary.senate.gov/imo/media/doc/2023-05-16_-_qfr_responses_-_altman.pdf)

<sup>89</sup> <https://www.forbes.com.au/news/innovation/canva-to-indemnify-users-for-what-its-ai-spits-out/>

<sup>90</sup> YouTube is figuring out its AI strategy by working with music labels.

Emilia David, The Verge, August 21, 2023

<https://www.theverge.com/2023/8/21/23840026/youtube-ai-music-copyright-monetization-universal>

<sup>91</sup> <https://www.alamy.com/blog/ethical-ai-image-initiative-alamy-announces-industry-partnership>

<sup>92</sup> <https://newsroom.gettyimages.com/en/getty-images/bria-partners-with-getty-images-to-transform-visual-content-through-responsible-ai>

<sup>93</sup> <https://www.musicbusinessworldwide.com/pink-floyd-david-gilmour-sony-music-debuts-generative-ai-project/>, <https://www.schumer.senate.gov/imo/media/doc/Dennis%20Kooker%20-%20Statement1.pdf>



- <sup>94</sup> Statement of Dennis Kooker President Global Digital Business & U.S. Sales, Sony Music Entertainment Senate AI Insight Forum: Transparency, Explainability, Intellectual Property & Copyright November 29, 2023 [https://www.schumer.senate.gov/imo/media/doc/Dennis\\_Kooker\\_-\\_Statement1.pdf](https://www.schumer.senate.gov/imo/media/doc/Dennis_Kooker_-_Statement1.pdf)
- <sup>95</sup> [https://commission.europa.eu/document/download/97e481fd-2dc3-412d-be4c-f152a8232961\\_en?filename=The%20future%20of%20European%20competitiveness%20\\_%20A%20comp%20etitiveness%20strategy%20for%20Europe.pdf](https://commission.europa.eu/document/download/97e481fd-2dc3-412d-be4c-f152a8232961_en?filename=The%20future%20of%20European%20competitiveness%20_%20A%20comp%20etitiveness%20strategy%20for%20Europe.pdf)
- <sup>96</sup> <https://www.wipo.int/web-publications/patent-landscape-report-generative-artificial-intelligence-genai/en/index.html>
- <sup>97</sup> <https://market.us/report/generative-ai-in-content-creation-market/> Market.us is not a company as notorious as KPMG, even if trusted by SACEM [https://www.goldmedia.com/fileadmin/goldmedia/Studie/2023/GEMA-SACEM\\_AI-and-Music/AI\\_and\\_Music\\_GEMA\\_SACEM\\_Goldmedia.pdf](https://www.goldmedia.com/fileadmin/goldmedia/Studie/2023/GEMA-SACEM_AI-and-Music/AI_and_Music_GEMA_SACEM_Goldmedia.pdf) but his survey is more useful for the needs of this analysis and is still consistent with that of KPMG <https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2023/generative-ai-survey.pdf>
- <sup>98</sup> [https://www.culture.gouv.fr/Media/medias-creation-rapide-ne-pas-supprimer/Chiffres-cles-2023\\_DEPS\\_Droits-d-auteur-et-droits-voisins\\_Fiche.pdf](https://www.culture.gouv.fr/Media/medias-creation-rapide-ne-pas-supprimer/Chiffres-cles-2023_DEPS_Droits-d-auteur-et-droits-voisins_Fiche.pdf)
- <sup>99</sup> Business Research Insights, 08/26/24 <https://www.businessresearchinsights.com/market-reports/ai-training-dataset-market-110110>
- <sup>100</sup> <https://openai.com/research/dall-e-2-pre-training-mitigations>
- <sup>101</sup> [https://cdn.prod.website-files.com/669550d38372f33552d2516e/66bc918b580467717e194940\\_The%20AI%20Risk%20Repository\\_13\\_8\\_2024.pdf](https://cdn.prod.website-files.com/669550d38372f33552d2516e/66bc918b580467717e194940_The%20AI%20Risk%20Repository_13_8_2024.pdf)
- <sup>102</sup> Content Owner Lawsuits Against AI Companies: Complete Updated Index. Audrey Schomer, Variety, 08/12/24 <https://variety.com/vip/content-owner-lawsuits-against-ai-companies-comprehensive-updated-index-1236101707/>
- Lawsuit says OpenAI violated US authors' copyrights to train AI chatbot. [Blake Brittain](https://www.reuters.com/legal/lawsuit-says-openai-violated-us-authors-copyrights-train-ai-chatbot-2023-06-29/). Reuters, June 29, 2023. <https://www.reuters.com/legal/lawsuit-says-openai-violated-us-authors-copyrights-train-ai-chatbot-2023-06-29/>
- 'New York Times' considers legal action against OpenAI as copyright tensions swirl. [Bobby Allyn](https://www.npr.org/2023/08/16/1194202562/new-york-times-considers-legal-action-against-openai-as-copyright-tensions-swirl). NPR, August 16, 2023. <https://www.npr.org/2023/08/16/1194202562/new-york-times-considers-legal-action-against-openai-as-copyright-tensions-swirl>
- At least 10 lawsuits have been filed this year against A.I. companies, accusing them of training their systems on artists' creative work without consent. [Sheera Frenkel](https://www.nytimes.com/2023/07/15/technology/artificial-intelligence-models-chat-data.html) and [Stuart A. Thompson](https://www.techtarget.com/whatis/feature/AI-lawsuits-explained-Whos-getting-sued), New York Times, 15 juillet 2023 <https://www.nytimes.com/2023/07/15/technology/artificial-intelligence-models-chat-data.html> <https://www.techtarget.com/whatis/feature/AI-lawsuits-explained-Whos-getting-sued>
- <sup>103</sup> <https://www.remotasks.com/en> , <https://crowdgen.com/> , <https://www.clickworker.com/> , <https://hivemicro.com> , <https://defined.ai/> , <https://help.neevo.ai/en/articles/3066177-how-when-how-much-will-i-be-paid> , <https://crowdsourcingweek.com/blog/top-5-crowdsourced-data-annotation-solutions/> , <https://www.washingtonpost.com/world/2023/08/28/scale-ai-remotasks-philippines-artificial-intelligence/>
- <sup>104</sup> <https://petapixel.com/2024/06/04/shutterstock-made-104-million-licensing-assets-to-ai-devs-last-year/>
- <sup>105</sup> <https://investor.shutterstock.com/static-files/11d9ca51-ceae-4756-854b-aa4912097a19>
- <sup>106</sup> <https://techcrunch.com/2024/06/25/backed-by-david-sacks-garry-tan-and-walter-isaacson-created-by-humans-helps-authors-license-their-creative-work-to-ai-models/?guccounter=1>
- <sup>107</sup> <https://authorsguild.org/advocacy/artificial-intelligence/ai-licensing-what-authors-should-know/>
- <sup>108</sup> <https://techcrunch.com/2024/06/08/deal-dive-human-native-ai-is-building-the-marketplace-for-ai-training-licensing-deals/>
- <sup>109</sup> <https://www.fairlytrained.org/> , <https://venturebeat.com/ai/the-first-fairly-trained-ai-large-language-model-is-here/>
- <sup>110</sup> [https://5a5ee099-3141-4217-af47-c61b445c2269.filesusr.com/ugd/6112c3\\_4c700dd417044c4aa268a4a4a9080c88.pdf](https://5a5ee099-3141-4217-af47-c61b445c2269.filesusr.com/ugd/6112c3_4c700dd417044c4aa268a4a4a9080c88.pdf)

<sup>111</sup> [https://www.lemonde.fr/economie/article/2024/09/14/les-contenus-utilises-pour-entraîner-les-ia-ont-une-valeur-et-un-prix\\_6317395\\_3234.html](https://www.lemonde.fr/economie/article/2024/09/14/les-contenus-utilises-pour-entraîner-les-ia-ont-une-valeur-et-un-prix_6317395_3234.html)

<sup>112</sup> <https://www.adami.fr/aide/4771/avec-quels-organismes-de-gestion-collective-ladami-a-signe-des-accords/>

113

[https://www.legifrance.gouv.fr/codes/article\\_lc/LEGIARTI000033677309#:~:text=Les%20%C3%A9%20%C3%A9ments%20d'information%20relatifs,d%C3%A9cret%20en%20Conseil%20d'Etat. ,](https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000033677309#:~:text=Les%20%C3%A9%20%C3%A9ments%20d'information%20relatifs,d%C3%A9cret%20en%20Conseil%20d'Etat.,)

[https://repertoire.sacem.fr/oeuvre-non-identifiee/conditions\\_generales](https://repertoire.sacem.fr/oeuvre-non-identifiee/conditions_generales) ,

[https://www.sacd.fr/sites/default/files/politique\\_generale\\_utilisation\\_sommes\\_non\\_repartissables.pdf](https://www.sacd.fr/sites/default/files/politique_generale_utilisation_sommes_non_repartissables.pdf)

<sup>114</sup> AI Act (108) : « En ce qui concerne l'obligation imposée aux fournisseurs de modèles d'IA à usage général de mettre en place une politique visant à respecter la législation de l'Union sur le droit d'auteur et de mettre à la disposition du public un résumé du contenu utilisé pour l'entraînement, le Bureau de l'IA devrait vérifier si le fournisseur a rempli cette obligation sans vérifier ou évaluer œuvre par œuvre les données d'entraînement en ce qui concerne le respect du droit d'auteur. »

Alexandra Bensamoun : « Il ne s'agit pas de tout lister œuvre par œuvre. »

[https://www.lemonde.fr/economie/article/2024/09/14/les-contenus-utilises-pour-entraîner-les-ia-ont-une-valeur-et-un-prix\\_6317395\\_3234.html](https://www.lemonde.fr/economie/article/2024/09/14/les-contenus-utilises-pour-entraîner-les-ia-ont-une-valeur-et-un-prix_6317395_3234.html)

<sup>115</sup> Alexandra Bensamoun : « L'important, c'est de sortir des postures et de reconnaître qu'il existe une chaîne de valeur. »

<sup>116</sup> Alexandra Bensamoun, à propos de la rémunération pour copie privée : « La loi décide de compenser un usage parce qu'il crée un préjudice économique. Il y a un prélèvement, ensuite réparti entre les titulaires de droits. Une commission pourrait fixer son montant. Certains titulaires de droits préfèrent cette rémunération automatique, parce qu'ils n'ont pas de force de négociation ou pas de grands catalogues de contenus jugés précieux pour l'IA. » [https://www.lemonde.fr/economie/article/2024/09/14/les-contenus-utilises-pour-entraîner-les-ia-ont-une-valeur-et-un-prix\\_6317395\\_3234.html](https://www.lemonde.fr/economie/article/2024/09/14/les-contenus-utilises-pour-entraîner-les-ia-ont-une-valeur-et-un-prix_6317395_3234.html)

<sup>117</sup> <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32006L0115&qid=1715051629968>

<sup>118</sup> <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32019L0790>

<sup>119</sup> Alexandra Bensamoun : « faut-il indiquer les sites sur lesquels on a trouvé les données et, si oui, avec quelle granularité ? Il ne s'agit pas de tout lister œuvre par œuvre, bien sûr, mais quelles informations faut-il retrouver pour permettre aux titulaires de droits de les exercer ? Car c'est le but clairement énoncé de la mesure. »

<sup>120</sup> AI Act (107) : « Tout en tenant dûment compte de la nécessité de protéger les secrets d'affaires et les informations commerciales confidentielles, ce résumé devrait être généralement complet en termes de contenu »

Alexandra Bensamoun : « l'AI Act prévoit que ce résumé des données soit mis à la disposition de tous. Cela aurait peut-être été intéressant de le confier à un intermédiaire [par exemple, le Bureau européen de l'IA] pour préserver, si nécessaire, la confidentialité. Cela étant, on ne peut pas invoquer le secret des affaires pour se dispenser d'appliquer la loi. Par ailleurs, ce concept est mentionné dans les considérants explicatifs [des annexes] du texte, mais pas dans la disposition elle-même, qui a force de loi. »

<sup>121</sup> Établi en 2018, le Code de bonnes pratiques contre la désinformation est un outil d'autorégulation, qui fixe les normes du secteur dans l'UE, sous la supervision de la Commission. Parmi les signataires du Code figurent Meta, Microsoft, Google, TikTok et Twitch

<https://www.euractiv.fr/section/application-de-la-loi/news/la-commission-europeenne-veut-se-doter-doutils-de-mise-en-oeuvre-du-dsa/>

<sup>122</sup> <https://authorsguild.org/advocacy/artificial-intelligence/ai-licensing-what-authors-should-know/>

<sup>123</sup> [https://www.linkedin.com/posts/cspla\\_report-aig-template-activity-7284973687169048578-YKor?utm\\_source=share&utm\\_medium=member\\_desktop](https://www.linkedin.com/posts/cspla_report-aig-template-activity-7284973687169048578-YKor?utm_source=share&utm_medium=member_desktop)

<sup>124</sup> <https://iswcnet.cisac.org/search> , <https://www.cisac.org/fr/services/services-dinformation/identifiants-internationaux>

---

<sup>125</sup> <https://web.isan.org/public/en/isan/0000-0000-3270-0000-H-0000-0000-N>

<sup>126</sup> <https://www.inpi.fr/programme-alliance-pi>

<sup>127</sup> [https://www.yacast.fr/mail/YACAST\\_Bilan\\_Radio\\_TV\\_2022.pdf](https://www.yacast.fr/mail/YACAST_Bilan_Radio_TV_2022.pdf)

<sup>128</sup> [https://www.spre.fr/wp-content/uploads/2021/06/SPRE\\_Rapport-de-transparence-FY20.pdf](https://www.spre.fr/wp-content/uploads/2021/06/SPRE_Rapport-de-transparence-FY20.pdf)

<sup>129</sup> <https://www.culture.gouv.fr/espace-documentation/Rapports/L-auteur-et-l-acte-de-creation> p.64

<sup>130</sup> <https://www.culture.gouv.fr/espace-documentation/Rapports/L-auteur-et-l-acte-de-creation> p.33

<sup>131</sup> [https://www.lemonde.fr/idees/article/2023/09/15/ia-generative-et-droits-d-auteur-la-culture-artificielle-ne-doit-pas-suivre-le-triste-chemin-de-la-malbouffe\\_6189539\\_3232.html](https://www.lemonde.fr/idees/article/2023/09/15/ia-generative-et-droits-d-auteur-la-culture-artificielle-ne-doit-pas-suivre-le-triste-chemin-de-la-malbouffe_6189539_3232.html)

<sup>132</sup> Ernst Fehr fait une évaluation par comparaison avec d'autres marchés publicitaires : « Si Google n'occupait pas une position de monopole dominant dans la recherche sur le Web, mais qu'il était confronté à une concurrence sérieuse en ligne, comme sur d'autres marchés publicitaires fonctionnels, 40 % de ce chiffre d'affaires constituerait une compensation marchande pour la contribution en valeur des médias »

[https://cominmag.ch/wp-content/uploads/2023/03/studie\\_leistungsschutz\\_fehradvice\\_170323-1.pdf](https://cominmag.ch/wp-content/uploads/2023/03/studie_leistungsschutz_fehradvice_170323-1.pdf)

<sup>133</sup> [Theses.hal.science/tel-01344612v1/document](https://theses.hal.science/tel-01344612v1/document)

<sup>134</sup> [https://www.lemonde.fr/idees/article/2015/06/19/le-triomphe-de-l-economie-de-l-utilite\\_4657975\\_3232.html](https://www.lemonde.fr/idees/article/2015/06/19/le-triomphe-de-l-economie-de-l-utilite_4657975_3232.html)

<sup>135</sup> [https://uploads.strikinglycdn.com/files/dal1f8c63-9083-4e8d-85cb-1554cf2c0ba2/2023-11\\_Controverse-rapport4.pdf](https://uploads.strikinglycdn.com/files/dal1f8c63-9083-4e8d-85cb-1554cf2c0ba2/2023-11_Controverse-rapport4.pdf)

<sup>136</sup> Rapport d'information de la commission des lois de l'Assemblée nationale déposé en conclusion des travaux de sa mission d'information sur les défis de l'intelligence artificielle générative en matière de protection des données personnelles et d'utilisation du contenu généré de Philippe Pradal et Stéphane Rambaud, rapport d'information n° 2207, 16e législature : [https://www.assemblee-nationale.fr/dyn/16/rapports/cion\\_lois/116b2207\\_rapport-information](https://www.assemblee-nationale.fr/dyn/16/rapports/cion_lois/116b2207_rapport-information)