Open:UK

State of Open: The UK in 2023
Phase Two, Part 2
"Show us the Money - AI Openness"

Index

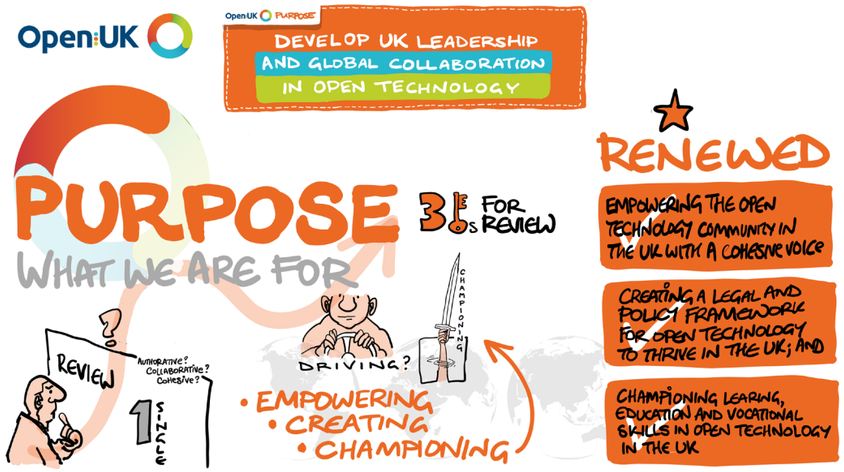
Cover photos are from our collaborator photographer Tiana Lea, who took 38 portraits for the OpenUK State of Open Exhibition sponsored by Arm[1]. We are grateful to Arm for their continued support of this work and will continue to document the people forming the UK Open Technology community in this way. Full list of cover participants at 7.

1 https://openuk.uk/photoexhibition/

PART ONE: INTRODUCTION

1.1 OpenUK as a unique organisation and its Reporting

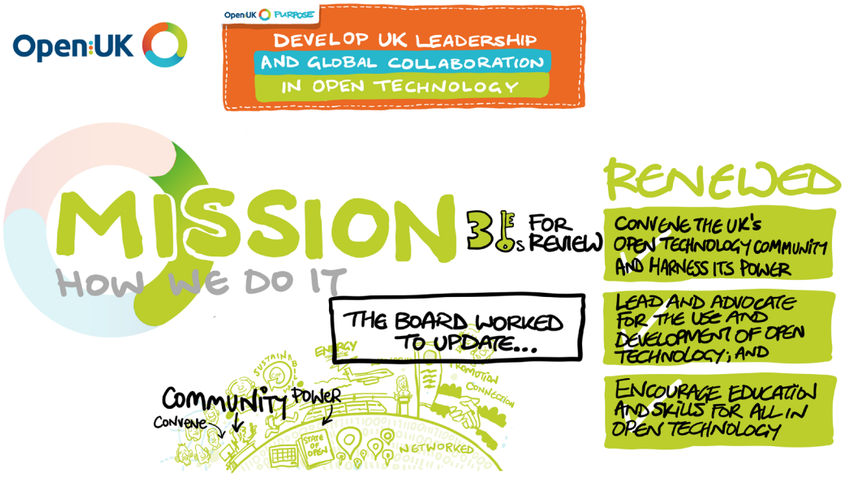
As the UK's organisation for the business of Open Technology - Open Source Software, open hardware and open data, implemented through open standards and open innovation our Purpose is UK Leadership and global collaboration in Open Technology.



OpenUK has a unique focus as a country organisation with this breadth of Open Technology. This arises partly from the breadth of its focus and recognition of the need to amalgamate the "Opens" in a future facing tech sector.

Also unique is the breadth of its activities which is not currently reflected in other country organisations in this space as can be seen in our Mission:

Community: Convene the UK's Open Technology community to harness its power;
Legal and Policy: Lead and advocate for the use and development of Open Technology; and
Learning: Encourage education and skills for all in Open Technology.



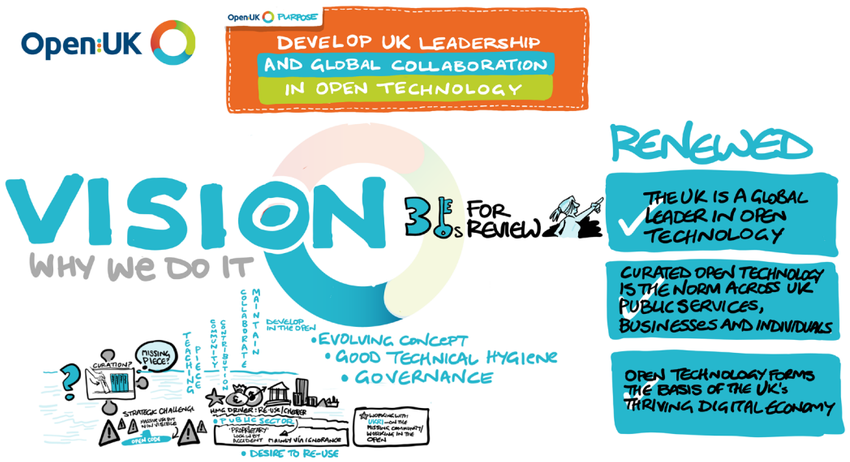
Our third unique factor is OpenUK's focus on people not companies. We are not a traditional membership organisation.

Of course our home grown companies - the normal focus of Open Source country organisations - are important to us and participate in our work. We also recognise the importance of the global companies employing much of our leading remote-working, UK-based, Open Technology workforce.

Our focus on people is important to us, to our sector and to a better understanding of it. People and their skills and talent are a fundamental basis of all technology. Our founders interviewed for this report reiterate and emphasise this through their interviews. They are not only the users but the creators of tech.

As is borne out by our research, there are a significant number of UK-based workers who form part of a global dispersed Open Source workforce, and an equally significant leadership in this space. This workforce, and their impact on the economy and UK tech sector is often missed. We will explore more on this, on migration and the future of work in Part 3.

Open Source has been the "Submarine under the Digital Economy" for many years. The internet, cloud, blockchain, AI, and importantly our national digital infrastructure are built on it. OpenUK's reporting allows this force to "up periscope" and share their value with you.



Within our Policy remit we have worked in partnership with Symmetry for 3 years, researching to build ground breaking reporting specific to Open Source Software[2].This research allows the boundaries to be pushed and the importance of the Values of "Open" to be clarified for all.

OpenUK is not a pay to play organisation and is funded by Sponsorship, Donations and Grants. We are grateful to our many sponsors and partners who make our work possible[3]. All are welcome to participate. Contact OpenUK on admin@openuk.uk.

---

2 "Open Source Software" is used throughout this report as a capitalised term to mean software where the source code is freely shared and the code is made available on an Open Source Initiative approved licence. However we recognise that this is more than a legal definition and the additional value of "Curated" Open Source Software - code with contributions, collaboration, community and good technical hygiene and governance all form a necessary part of "Open Source Software."

3 https://openuk.uk/participants/sponsors/

1.2 Introduction: Show me the money Openness and AI

Amanda Brock
CEO
OpenUK

This State of Open: The UK in 2023, Part 2, “Show us the Money - Openness and AI” has been a somewhat difficult introduction and Report to write with even its name changing from “Open Source AI” to “AI Openness.” Difficult, not simply in keeping up with the pace of change and challenges to our views on AI and “open”, from a technical, business and societal perspective, but also due to a relentless pace of development and ensuing information overload.

As with other contributors to the Report, this is not my first rodeo. I have previously participated in critical leaps in technology and their impact. A lawyer during the early stages of the dotcom boom, employed in a leading UK Internet Service Provider, I was one of the first with understanding of the technical, practical, business and human challenges of the internet. I participated in the process of creating internet regulation in the early 2000’s. Open Source Software brought a second opportunity. AI creates the third lens I have observed this process through. A privilege to have engaged in the evolution of three of the most important technologies and their regulation - the internet; Open Source Software; and AI. Possibly the most transformational technologies of our time.

A short history lesson

Coding began as a collaborative process. Groups of coders, primarily in universities collaborated and shared. Legislative decisions to extend copyright law to code set the direction of our technology futures. That direction was one of IP ownership and restrictions; of licence-based usage. Consequently the propensity for collaboration and sharing was reduced, and ultimately innovation hindered. This directly led to code, and ultimately technology essential to society today, being in the hands of a few. IP-harvesting for new technology spaces is often compared to an “arms race.” Those holding IP may defend off new market entrants and innovators.

Open Source Software, fundamentally creates the ability to freely share and build on others’ work by recycling and reusing their code and the IP within it - “standing on the shoulders of giants.” This facilitates innovation - improved by collaboration - removes lock-in, facilitates interoperability, establishes open and de facto standards, and enables trust through open and transparent deliverables and processes. Perhaps it can be considered a correction of an early-stage legislative decision and consequential IP-dominated tech evolution. Open Source Software, As the choice of the code developers, may be seen as a societal decision by those working and creating code, to counter the IP restrictions of legislators in previous decades.

Lessons Learned

Whether you agree that Open Source Software is a societal correction or not, the impact of IP and regulation and knock-on-consequences are undeniable. The impact of copyright and other IP on past software development and their consequences provide a clear lesson to today’s decision makers.

Regulators are struggling to manage the update of our internet laws - many of those laws are the same ones I helped establish almost 25 years ago. Whilst appropriate at the time, they are simply not fit for the unimaginable world our internet has become today. We can be certain AI will also create a further unimaginable world in future. Those making policies, regulation and regulatory processes must learn from recent history to ensure their decisions have a positive impact on that unimaginable world to come.

Governments are considering the inevitable policies and regulation: both deciding the level of regulation required - the prescriptive and controlling European Commission approach, both loved and hated by those contributing Thought Leadership and case studies to this Report, versus a more agile, principles based, approach allowing the flow of innovation; and deciding who will own our technology futures. I use the word “our” advisedly. These technologies will impact all of us.

The fundamental decision to be made is a simple one. Will our government’s approach be one of control established through transparency and building trust engendered by openness, or will it be a decision to seek control by closing down innovation, thereby leaving the outputs of this most important of innovations, in a proprietary “black box” without transparency? Clearly the latter would fail to either learn from the past technology decisions or to engender trust. It would also leave critical technology in the hands of the few.

What is AI Openness

Of course there are shades of openness that may apply to our discussion on AI. It has become clear that we need a tautology for AI and openness to provide clarity. As our case studies and Thought Leadership came together for the Report, it transpired that the varying aspects of “openness” represented by the breadth of Open Technology OpenUK represents -Open Source Software, open hardware and open data, enabled by open innovation and open standards - must be considered across AI. This Part 2 of our Report creates a body of work and emerging critical opinion on the Openness of AI for the first time. With contributions from UK organisations like The Tony Blair Institute and Turing Institute and global leaders in Open Technology it sets the groundwork for understanding. Openness will, of course, democratise AI technology.

Our thought process starts with what is AI openness? What aspects of AI might be opened up and how? What are the consequences - both benefits and risks - of opening it up or not? This necessitates a new approach for this Part of our Report. Rather than purely focus as we have in Part 1 and in past reporting, on Open Source Software, for the purpose of this Part 2 of our Report, we take a holistic view and consider Open Technology, and openness adding open data, open standards and open innovation, to lead the conversation in this space.

What I know for sure is that here in the UK we have leadership in both Open Technology and AI which uniquely position the UK to lead in this space and that leadership will lead the UK to the money.

PART TWO: STATE OF UK OPEN AI

2.1 Runa Capital GitHub Landscape and AI

Using GitHub's API we automatically collected data on all Open Source repositories with more than 1,000 GitHub stars at the end of June 2023. Then our investment team manually re-viewed top repositories, and detected UK tech businesses among their owners.

We also checked which top repositories mainly contain software code related to machine learning, neural networks, large language models, computer vision and other AI topics. It al-lowed us to find top-5 AI software repos by GitHub stars, owned by UK tech companies.

Top AI Software Repos (owned by UK Tech Companies)

	Repo	Description	GitHub Org	UK Entity	Created	Repo Stars(K)
1	Significant-Gravitas/Auto-GPT	Experimental autonomous version of GPT-4	Significant-Gravitas	SIGNIFICANT GRAVITAS LTD	2023	141.3
2	Stability-AI/stablediffusion	Deep learning-based text-to-image model	Stability-AI	STABILITY AI LTD	2022	25.3
3	mindsdb/mindsdb	ML/AI layer for databases	mindsdb	MINDSDB LTD	2018	17.0
4	deepmind/deepmind-research	Code for DeepMind's publication about AI	deepmind	DEEPMIND TECHNOLOGIES LIMIT	2019	11.9
5	unifyai/ivy	ML acceleration framework	unifyai	DEEP IVY LTD.	2021	11.5

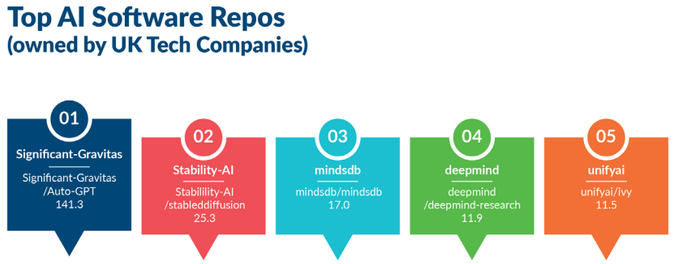


Figure 1. Top AI Software Repos (owned by UK Tech Companies)
Source: Runa Capital analysis of GitHub API Data

But when you get to the heart of what AI is and how it can be applied to unlock value in businesses and everyday life, you have to admit that we're standing on the edge of a revolution. This revolution is likely to change our lives significantly in the short term, and perhaps tremendously so in the medium term. Matt Barker, Founder, Jetstack

The big 'step changes' in technology - like 'the internet' then 'Open Source Software / Linux', and then 'public cloud' - have always carried with them a reduced barrier to entry for start-ups. I consider AI to be one of these new step changes. Where it might have taken a team of 5 and $5m 10 years ago, AI might create opportuni-ties for a couple of smart people and a very small amount of money to get off the ground. Most big software companies will be worried about the ability to be out-competed by AI startups, the increased competition will lead to big winners and big losers. But what's clear is that a team taking advantage of Open Source is going to be one of the best ways of making yourself successful with the technology.
Matt Barker, Founder, Jetstack

2.2 Information on AI from the survey

2.2.1 Office of National Statistics

Across the UK, businesses, consumers and the government are engaging in dialogue with the scientific and technology communities to better understand the massive impact of AI, and par-ticularly Generative AI, on businesses and everyday life.

So much so that the UK Office for National Statistics (ONS)[4] recently published the results of an opinion survey on AI - a quick-turnover bellwether of public perceptions.

The vast majority of responses indicates that the public remains on the fence about the impact of AI on the UK, possibly due to the increased uncertainty about its capabilities.

What is encouraging is that 53% of the population surveyed by the ONS felt confident to give a partial explanation of what AI is, and only 9% have never heard anything about AI.

It is not surprising that more than half the population can broadly explain what AI is, as the UK is one of the largest contributors to AI projects globally.

ONS Survey: Have you ever heard of the term Artificial Intelligence (AI)?

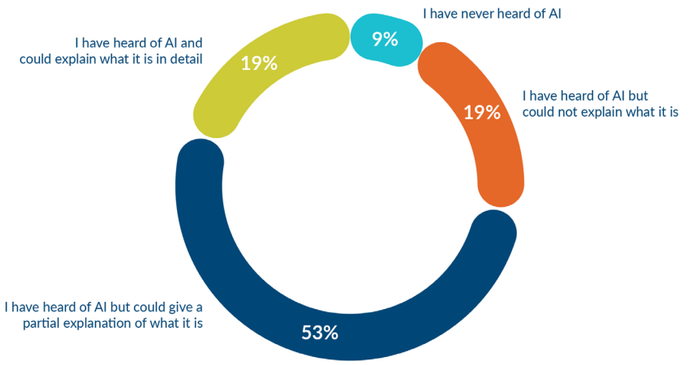


Figure 2. ONS Survey: Have you ever heard of the term Artificial Intelligence (AI)?
Source:UK Office for National Statistics public opinions survey (4-14 May 2023)

The results are telling about what the public believes, especially in terms of impact[5].

ONS Survey: What impact do you think AI will have on the UK?

Figure 3. ONS Survey: What impact do you think AI will have on the UK?
Source: UK Office for National Statistics public opinions survey (4-14 May 2023)

4 ONS. (2023). Artificial Intelligence (AI) awareness, use and impact, Great Britain. UK Office for National Statistics. Retrieved from: https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/datasets/artificialintelligenceaiawarenessuseandimpactgreatbritain
5 ONS. (2023). Public opinions and social trends, Great Britain: 4 to 14 May 2023. Office for National Statistics. Retrieved from: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/bulletins/publicopinionsandsocialtrendsgreatbritain/4to14may2023

Regardless of what happens, I'm sitting here thankful that I can be a part of yet another tech movement. For those that worry about the consequences to the human race, I empathise, but I also know that the genie is already out of the bottle. At this point, it's better to embrace and help direct it to make it useful and safe, than it is to try and fight the tide. I, for one, am ready to get going.
Matt Barker, Founder, Jetstack

2.2.2 Organisation of Economic Cooperation and Development

The Organisation of Economic Cooperation and Development (OECD), using data by GitHub[6] found that the UK is the third largest contributor to public AI projects (of all levels of impact), accounting for 4% in 2022. The top countries in AI in 2022 were India (23%), and the entire EU27 countries collectively, and the USA each contributing 14%. The UK in 2023 has so far received $29.2 billion in venture capital investment for AI according to the OECD[7], approximately 8.8% more than 2022.

Comparatively, the USA has received so far $461.4 billion, 20.7% more than in 2022. The vast difference in venture capital investment for AI between the UK and the US illustrates that the UK still has some way to go to become a global leader in AI.

OECD Contributions to public AI projects by country, 2011-2022

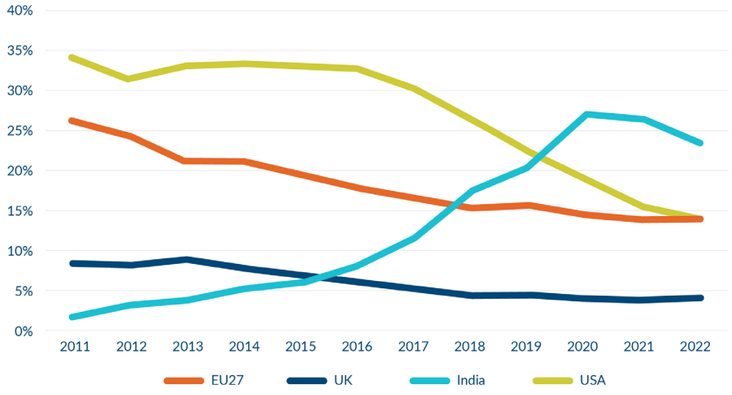


Figure 4. OECD Contributions to public AI projects by country, 2011-2022
Source: OECD 2023

6 OECD.AI (2023). Visualisations powered by JSI using data from Github, accessed on 27/6/2023. Retrieved from: www.oecd.ai . Supported by the Patrick J. McGovern Foundation.

7 OECD.AI (2023). Visualisations powered by JSI using data from Github, accessed on 27/6/2023. Retrieved from: www.oecd.ai . Supported by the Patrick J. McGovern Foundation

2.2.3 OpenUK Survey

To further this conversation, OpenUK included three challenging, forward-thinking questions pertinent to the relationship between Open Source Software and AI in our proximity survey distributed in June 2023[8]. While the sample surveyed is not representative of the UK public, as most of the respondents are from industries where AI has been embedded early on in the business, the findings are noteworthy.

"There are two approaches we're seeing to the regulation of AI — the top down, big bang approach that the EU is taking which is complex, multifaceted, and in some respects quite prescriptive. On the other hand, we've got an innovation-focused approach that the UK and the US are taking; devolving responsibility to sectoral regulators and federal agencies. A justification given for that is because the sectoral regulators have the expertise on the use cases, which is where the rubber hits the road." Chris Eastham, Partner, Fieldfisher

The survey asked if AI ownership concerns can be solved by Open Source Software

Open Source Software offers the solution to concerns about AI ownership

Figure 5. Open source software offers the solution to concerns about AI ownership
Source: Q38 OpenUK State of Open Survey 2023

The largest proportion of respondents (45%) are neutral, with 40% agreeing and 15% disagreeing.

The large percentage of those that have not firmly decided indicates that there is a lot of uncertainty and complexity around AI ownership.

"I think Open Source does have a substantial part to play in democratisation. I wouldn't like to see all of AI sitting within the few big tech companies. That would be a bad thing."
Chris Eastham, Partner, Fieldfisher

8 Please refer to the methodology for description of the survey and see The State of Open Source: The UK in 2023 Phase Two Part One for more survey outputs, detail and questions.

The question of copyright and AI is a highly contentious one

Only 14% of survey respondents believe that AI should have copyright in the code it creates. In fact, the majority of respondents disagree (45%), but a significant proportion seem to be undecided (41%).

AI should have copyright in the code it creates

Figure 6. AI should have copyright in the code it creates
Source: Q39 OpenUK State of Open Survey 2023

The latter group will be critical in informing and shaping the discussion on what is capable of owning copyright, as creativity now depends not only on conventional capabilities of humans.

"If you are publishing code that is subject to Open Source licence terms, then you need to follow the terms in the relevant Open Source licence. That often means attribution, at the very least." Chris Eastham, Partner, Fieldfisher

This topic is explored among the interviewees in the next sections of this Report.

Time and time again we have seen the Open Source model applied to drive outstanding speed, scale, and innovation. Yes it was partly thanks to a leak, but Open Source has already replicated 96% of OpenAI's capabilities in a matter of months using LLaMA. Can one company be confident in its ability to compete against millions of developers working towards the same goal, in the open and with the greater visibility of bugs and issues that affords? I'm certainly not saying there won't be a place for proprietary models that take advantage of specific requirements, but I'm not going to bet against the power of the Open Source model. On top of this, Open Source code itself is also going to get its own efficiency boost thanks to the application of AI to help build and optimise it! Just look at the power of applying Co-Pilot by GitHub to your coding practices, or K8sGPT to your Kubernetes cluster. This is only just the beginning. Matt Barker, Founder Jetstack

The final question relates to transparency

Datasets used to train should be disclosed in SBOM

Figure 7. Datasets used to train should be disclosed in SBOM
Source: Q40 OpenUK State of Open Survey 2023

There is consensus among respondents for the disclosure of datasets that have been used to train AI in the software bill of materials (SBOM), with 85% of them agreeing that data sets should be disclosed in a Bill of Materials. 13% remain neutral and only 2% disagree.

This is an issue that affects the supply chain of AI production, but at the same time it can relate to commercial confidentiality, data privacy, ethics and, of course, the question of attribution.

"The people who set up the AI systems, when an anomalous result is produced, are incapable of telling you how that came about. So it really is irrelevant whether a) the algorithms and b) the source of the training material was conventional copywritten Open Source, because the problem is not one of where you draw the information from. How you process the problem is one of explicability. And it's inherent in the nature of statistical predictions that you cannot determine how they actually work. They are black boxes. The answer to the question of transparency is that there is no easy answer. And beware of anybody who tells you there is an easy answer." Iain Mitchell, Honorary KC, OpenUK

PART THREE: OPEN AI THOUGHT LEADERSHIP

3.1 The 2023 AI Landscape

3.1.1 Pace of innovation

The pace of AI innovation in the first few months of 2023 has been unprecedented. The decision to include a small update on AI in OpenUK's 2023 Phase Two Report, made at the beginning of the year has today necessitated dense reporting and our splitting of Phase Two's "Show US the Money" Report into two: Part 1, "The Economics of Open Source Software," made up of economic analysis and a review of UK Open Source businesses and Part 2, "AI OpennessI" to give each subject the appropriate attention it deserves. Part 1 is released contemporaneously with this Part 2 on 13 July, 2023.

The first six months of 2023
On 13 March 2023, the world was abuzz with talk of the company "OpenAI" and it's Generative Pre-trained Transformer, ("ChatGPT") release 4 ("GPT-4") launch[9]. This AI Chat bot tool - effectively a neural language calculator that has been seen by many as being game-changing in the AI space. The term "generative AI" has entered common parlance and the debate as to whether AI will destroy humanity began. At the same time the company name "OpenAI" has led to confusion. According to Elon Musk, an initial investor in Open AI, "OpenAI was created as an open source (which is why I named it "Open" AI), non-profit company to serve as a counterweight to Google, but now it has become a closed source, maximum-profit company effectively controlled by Microsoft". OpenAI is however very much a for profit company and its ChatGPT is not open.

The Open Source community was operating at a disadvantage in early 2023. When it came to the development of AI, they lacked access to an LLM - Large Language Model. The vast cost of creating these models had put the possibility of the creation of one out of reach of open developers. The LLMs have very much been the "secret sauce," carefully protected by large companies which can afford to create them and the potential centre of very lucrative ecosystems. So vase are their cost that 90% of their compute power is said to sit with 7 companies in the US and China.

Access to an LLM came about however via the leak of Meta's"LLaMA" LLM. Meta had shared[11] the LLM with researchers on a restricted licence, made available strictly for research purposes. LLaMA was provided without instruction or conversation training and did not have reinforcement learning from human feedback (RLHF)[12]. Explained by Meta as "a state-of-the-art foundational large language model designed to help researchers advance their work in this subfield of AI."

However, on 8 March, press began reporting[13] that the LLaMA LLM had been leaked to the wider open development community and was effectively now freely available.

9 https://openai.com/gpt-4
10 https://twitter.com/elonmusk/status/1626516035863212034?lang=en
11 MetaAI. (2023). Introducing LLaMA: A foundational, 65-billion-parameter large language model. Retrieved from: https://ai.facebook.com/blog/large-language-model-llama-meta-ai/
12 Wikipedia. (2023). Reinforcement learning from human feedback. Retrieved from: https://en.wikipedia.org/wiki/Reinforcement_learning_from_human_feedback
13 https://www.theverge.com/2023/3/8/23629362/meta-ai-language-model-llama-leak-online-misuse

Within a month of the "escape" of LLaMA into the wild, individuals and groups of individuals experimenting with this had created instructions tuning, quantization, quality improvement, human evals, multimodality, RLHF."

The commercial response
"We have no moat" says a Google employee in an internal Memo leaked in May[14], "we aren't positioned to win this arms race and neither is OpenAI," acknowledging the power of Open Source development. While we've been squabbling, a third faction has been quietly eating our lunch. I'm talking, of course, about Open Source. Plainly put, they are lapping us. Things we consider "major open problems" are solved and in people's hands today."

Not the opinion of Google, but simply of a single Google employee in an internal response to the inevitability of AI in 2023 - "the leaked memo". The use cases shared by this employee demonstrates the accessibility of AI to individuals and the ability of developers to experiment and develop AI on their own, examples include running a LLM on a personal mobile phone and being able to finetune a personalised AI on a personal laptop in an evening.

The memo goes on to discuss business models and the fact that those commercial entities included in the "arms race" have no secret sauce to protect their innovation and lock in economic benefit to their AI. With the open community access to LLaMA, the inhibitors to open innovation were gone.

The Google memo called out the concern that companies might not have access to a traditional route to commercialisation or revenue generation. Customers simply won't pay in future for the finesse that can be offered in an expensive proprietary offering from a company when an "almost-equivalent" is freely and openly available.

This highlights with crystal clarity the pace of development of "Generative AI".

3.1.2 The "open" Outputs

All of this hype has unsurprisingly caused an explosion of interest in AI and a boom in AI companies reminiscent of the early days of the dot com era.

Alumni from Deepmind and Meta Open Source solutions for enterprises has seen a 4 week old company raise $100m[15], from the US.

Here in the UK companies like Sustainability AI and Significant Gravitas shot to stardom. As the creator of Auto-GPT built on OpenAI's GPT-4, Significant Gravitas' repo is now number one in the U K and the second AI project to gain 100,000 stars on GitHub and is the number one code repository in the UK as can be seen from the Runa Capital data analysing UK GitHub repositories at section 4.1. It is an experimental Open Source Python Application released on 30 March, using the Chat GPT Open API. Founded by game developer Toran Bruce Richards, Significant Gravitas is a company registered in the Scottish town of Oban and Toran is based

14 https://www.semianalysis.com/p/google-we-have-no-moat-and-neither
15 TechCrunch. (2023). France's Mistral AI blows in with a $113M seed round at a $260M valuation to take on OpenAI. Retrieved from: https://techcrunch.com/2023/06/13/frances-mistral-ai-blows-in-with-a-113m-seed-round-at-a-260m-valuation-to-take-on-openai/

in Edinburgh. Auto-GPT is remarkable in its ability to act autonomously with minimal human intervention and has an ability to self-prompt. When it is provided with an end goal, it will self prompt each task required to reach that point. It does not live on a browser and requires specific software to be downloaded by the user. "This application has the potential to revolutionise the way we approach complex tasks that require human-level intellect. It is a promising step towards artificial general intelligence (AGI)."[16]

The opening of AI will only see more small companies and innovators developing ground-breaking innovation and world-leading AI here in the UK and globally.

3.1.3 The Legal Questions

3.1.3.1 AI, Risk and Regulation

As we look to regulation we of course resort, as with Part 1 of this Report to questions of risk and understanding of it. In an emerging area such as AI risk will take time to understand. To create trust requires transparency, through transparency can also come control.

On the 11 July, the FT reported[17] Nick Clegg, former leader of the UK's Liberal Democrat party and now Head of Global Affairs at Meta, "Openness on AI is the way forward for tech: The case for transparency is growing as the best way to combat fears on the developing technology. In acknowledging that risk cannot be eliminated he recognised how it could be mitigated propos-ing 4 steps: 1. Transparency as to how a system works; 2. Accompanying openness with collaboration; 3. AI systems should be stress tested; and 4. Companies ought to share details of work as it develops whether through papers, research etc or sharing the work itself.
Regulation to meet the needs of risk mitigation must be implemented in an agile way and have the flexibility to evolve over time. A principles-based approach regulating the use case of the 'utilisation of the tool' and not the tool itself (in the main) makes sense. The onus lies with the user of the AI and puts responsibility back on them to exercise discernment in their selection of AI and management of its use appropriate to the context of that scenario. This means an appropriate assessment and understanding of the particular tool and the data sets on which it was trained may be required. Where the context is a regulated sector a different, possibly greater, degree of care will be required than in other utilisations.

16 https://www.itworldcanada.com/post/significant-gravitas-launches-auto-gpt
17 https://www.ft.com/content/ac3b585a-ce50-43d1-b71d-14dfe6dce999

3.1.3.2 Data

The Inputs
AI is trained on data sets. This raises a number of issues regarding whether: there ought to be a requirement to disclose the data upon which the AI is trained; whether that should be a form of bill of material, or whether a higher level disclosure such as a data card would be adequate; whether publicly available data may be used to train AI without breaching copyright; and the rights of an individual to know what data about them is being used to train an AI.

The outputs
As well as the right to openness in sharing what data AI is trained on to allow an understanding of behaviour such as bias, there is also a need for understanding of "automated" decisions made by AI which may impact an individual. Data rights in the UK are discussed in more detail at 3.4.

3.1.3.3.Intellectual Property

Inputs
The content of the data on which AI is trained may include copyright works or a part thereof. This may also include software under a licence as we have seen in the case of Open Source Software used by tools like CoPilot. Issues around the need for attribution and other licence terms being complied with are very real concerns for the creators of the work being gobbled up in the AI process.

Outputs
Whether AI is to be granted the right to ownership of intellectual property in its creations remains to be seen and will be a question for laws and courts. In March the US copyright office released guidance[18] saying that, U.S. copyright law requires "human authorship" to register copyrighted works, because "author" as used in the Copyright Act, "excludes non-humans."

A consideration of copyright in the work created by it and who the owner of the copyright in that new work generated by the AI might be appears likely to depend on the way the generative AI is used and whether the data set is provided with it under licence or created by the questioner and that may well create the difference between the questionnaire and the creator owning it. This is another matter for regulation.
Whether this belongs to the creator of the AI or the asker of the question remains to be seen but likely the latter. To some extent this will depend on who compiles the data set on which the question is answered.

On 11 July 2023, "ChatGPT sued for copyright infringement," "57. Because the OpenAI Language Models cannot function without the expressive information extracted from Plaintiffs' works (and others) and retained inside them, the OpenAI Language Models are themselves infringing derivative works, made without Plaintiffs' permission and in violation of their exclusive rights under the Copyright Act."[19] Others have already spent considerable time delving deep into the legal weeds around this and similar issues[20] and that detail is not the point.

18 https://www.copyright.gov/ai/
19 https://www.courtlistener.com/docket/67569254/silverman-v-openai-inc/
20 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4464001

As well as a consideration of ownership of rights in any outputs, there is a question of liability in these. Will that be the same person as the owner of the rights? Interestingly this may not be the case.

3.1.3.4 Definitions: What are Open AI, Open Source AI and AI openness?

An ontology: Terminology like "open AI," "Open Source AI" and "open AI innovation" are being regularly used without clarity at the present time. Work is under way to consult and create more clarity in this space. There is a clear need for an agreed ontology to support regulation.

Open Source Initiative: Deep Dive AI was launched in 2022, as a podcast that culminated with a Report. The OSI has extended this into a Consultation on the meaning of "Open Source AI" and stakeholders are being asked to submit to a global drafting process to create a definition of "Open Source AI" being Open Source as applied to AI/ML. A series of consultations and a Call For Proposals/.

Open Knowledge Foundation[22]: Consultation underway on open AI.

The Role of Open Data Platforms : Industry has spent time already working on - access to data will be critical and we need access to this to be regulated. Well managed open data. Clear regulation is needed to create certainty in the market and to make space for the good actors. Needs a framework around AI regulation and a framework for data collection.

3.1.3.5 Standards

AI will require standard setting at pace and the more open and interoperable these are the more they will facilitate innovation.

21 https://opensource.org/deepdive/
22 Open Knowledge Foundation. (2023). Debriefing the Open Definition workshop at RightsCon. Retrieved from: https://blog.okfn.org/2023/06/26/debriefing-the-open-definition-workshop-at-rightscon/

3.1.4 Benefits and Risks of Openness in AI

AI and the UK
Whilst the apparent "arms race" has been playing out in the US, here in the UK? Prime Minister Sunak announced the autumn AI Summit[23] during London Tech Week in early June.

UK White Paper on AI: The UK Office of AI White Paper consultation began in March and closed in June. A response is expected in September.

UK Task Force: The UK, led by the Prime Minister Rishi Sunak, set up a Foundation Modal Taskforce led by Ian Hogarth.

[24]**UK AI Standards Hub**: Role of technology standards are changing, because of changing role in society therefore need to be made more accessible - AI Standards Hub is leading the way on this in the UK.

PM's Global AI Summit: Whilst the UK may not have a moat either, it is an Island nation, with a currently huge focus on AI.

Perhaps the first sign of recognition of change in the political and regulatory landscape in the UK for OpenUK, was an invitation to our CEO from Lord Camrose, the UK AI Minister, to attend a small Round Table held by him during London Tech Week. This participation along with a recent OpenUK Roundtable with the UK Office of AI, enable OpenUK to lead the discussion on openness of AI in the UK, through contribution to the consultation process are clear indications of a realisation that Open Source Software, open data and other aspects of open innovation are going to be critical to the future of AI and its inevitable regulation.

23 Open Knowledge Foundation. (2023). Debriefing the Open Definition workshop at RightsCon. Retrieved from: https://blog.okfn.org/2023/06/26/debriefing-the-open-definition-workshop-at-rightscon/
24 Gov.UK. (2023). Tech entrepreneur Ian Hogarth to lead UK's AI Foundation Model Taskforce. Retrieved from: https://www.gov.uk/government/news/tech-entrepreneur-ian-hogarth-to-lead-uks-ai-foundation-model-taskforce

3.2 Thought Leadership: "Open Source AI" Definition

Stefano Maffulli
Executive Director
The Open Source Initiative

The Open Source communities have relied for decades on copyright and its sibling, copyleft[25] but this approach is showing its limitations[26] with modern technologies. Artificial Intelligence and Machine Learning are posing an even larger challenge to the principles of Open Source than cloud and mobile did.

To start, AI and ML blend the boundaries between software and data. The AI systems introduce new artefacts for which the applicability of copyright law is questionable. The Generative AI systems also pose new and intricate legal challenges to the many established understandings of patents and trade secrets. And the large quantities of data required to build functional ML systems also attract other laws – from privacy protection to security to non-discrimination and accessibility laws – all the way to basic human rights protections. Much of the legal principles at the base of OSI Approved Licences are already being challenged in these contexts.

> Despite the popularity of the term "Open Source AI," it has no shared and agreed definition. And despite the popularity of software licences applied to ML models, not everybody agrees about the applicability of their terms.

To bring the principles of Open Source with us for the tech that comes next, we must think hard, carefully and quickly about how to adapt the guiding principles of "open" to the AI/ML field. OSI is calling for stakeholders to join the global drafting process of a definition of "Open Source AI" being Open Source as applied to AI/ML.

Looking for a solution

Deep Dive: AI is a series of events organised by the Open Source Initiative (OSI) as a wide consultation with communities of practice, researchers, experts of ethics and human rights to explore the challenges and opportunities of Open Source AI. First produced in 2022, the 2023 series consists of three parts:

- A webinar series featuring experts on AI and Open Source, airing between September and October (the call for speakers is open until August 4th):
- An in-person session at All Things Open on October 16th; and
- A report summarising the findings of the 2023 Deep Dive.

25 https://www.gnu.org/licenses/copyleft.en.html
26 Maffulli, S. (2023). OSI Opinion. The importance of Open Source AI and the challenges of liberating data. Retrieved from: https://blog.opensource.org/the-importance-of-open-source-ai-and-the-challenges-of-liberating-data/

The goal of Deep Dive: AI is to help the Open Source community understand the implications of AI for Open Source. The series addresses a range of topics focused on the implications of applying Open Source principles in the development and use of AI models

Deep Dive: AI is a timely and important initiative. As AI becomes more pervasive, it is essential that the Open Source community promote the development of AI systems that are used in an ethical manner, consistent with the ethos of Open Source. The OSI is soliciting input from the community throughout the series. The event is a valuable resource for anyone interested in the future of Open Source AI, especially with regard to licensing issues and the very definition of Open Source in an AI world.

The webinars will be particularly informative, and the session at All Things Open will offer a diverse range of perspectives on the future of AI and Open Source.

Focusing on Open Source principles

> As AI technologies become more prevalent and influential, Open Source and AI developers are looking for licensing models that promote innovation while limiting harm. The emergence of AI has raised complex questions regarding ownership, accountability, and fairness. Developers and organisations utilising AI technologies are increasingly recognising the need to align their work with ethical principles and societal values.

Deep Dive: AI will address one of the primary challenges in this new reality, providing frameworks to promote the responsible use of AI with Open Source licensing, for example, curtailing malicious or harmful applications. The concerns that AI technologies can too easily violate privacy, enable surveillance, or facilitate discrimination are at the forefront of researchers and civil rights advocates, highlighting a tension with some of the core principles of Open Source.

Another aspect to be addressed is the protection of data used in AI models. Open Source licences are being updated to specify how data collected and processed by AI systems should be handled, ensuring compliance with privacy regulations and ethical data usage. Additionally, there is growing awareness around the potential risks associated with AI-generated content, such as deepfakes or misinformation.

Deep Dive: AI[27] will explore collaboration models that incentivise contributions to AI projects while safeguarding the open nature of the systems.We encourage everyone concerned about the intersection of AI and Open Source to consider how they might update their licences to address the unique challenges posed by AI. Deep Dive: AI is here to help, exploring responsible use, data protection, content transparency, and community collaboration and providing insights to help evolve licences that shape AI development in a manner aligned with societal values.

27 OSI. (2023). Defining Open Source AI. Retrieved from: https://opensource.org/deepdive/

3.3 Thought Leadership: AI and Open: where are we now?

TIDELIFT

Luis Villa
Founder Tidelift and Open(ish) Machine Learning News

I've been trying to closely follow the nexus of open and AI for about six months now, through my newsletter at openml.fyi.[28] This is simultaneously the most exciting, and the most complex, thing to hit 'open' in two decades — so there has been a lot to digest! In mid-June I put together a lengthy summary of the past six months of news[29], to help people analyse where open AI is and where it could go. Since that is 4,000+ words, I'm excited to pull out a few of the most important parts for this report

Open AI is not yet fully defined

No major open organisation or organised movement has formally defined "open AI". However, that could change soon: the UK-based Open Knowledge Foundation is working on a a definition[30], and the Open Source Initiative is also working on a definition[31]. Nevertheless, the term is being used (and abused). Many independent developers are working on AI "in the open"—by using open tools, open collaboration, and open data.

Open AI is plausible

> Just a few years ago, I would have said that meaningfully open AI would never happen, because of the high cost of training any large model would make distributed development impossible. That has significantly changed.

All block quotes below are from my openml.fyi "state of open" summary[32].

[T]he cost of training is dropping (from ten of millions of dollars to hundred of thousand even for large models)[33] and new techniques like Low-Rank Adaptation (LoRA) and QLoRA are making some types of training possible on desktop machines. Continued interest in lower-resource training from academics, hobbyists, and non-FAANG companies will likely accelerate creation of tools and techniques that enable distributed model development.

Open AI is broad

Critically, this moment is not just about Open Source Software. Open AI also implicates open culture, like Wikipedia, and many values that have traditionally paralleled, but not been formally associated with, Open Source Software, like digital privacy and antitrust law.

28 https://www.openml.fyi/
29 OpenML. (2023). Taking stock of open(ish) machine learning / 2023-06-15. Retrieved from: https://www.openml.fyi/taking-stock-23/
30 Open Knowledge Foundation. (2023). Debriefing the Open Definition workshop at RightsCon. Retrieved from: https://blog.okfn.org/2023/06/26/debriefing-the-open-definition-workshop-at-rightscon/
31 OSI Blog. (2023). Now is the time to define Open Source AI. Retrieved from: https://blog.opensource.org/now-is-the-time-to-define-open-source-ai/
32 OpenML. (2023). Taking stock of open(ish) machine learning / 2023-06-15. Retrieved from: https://www.openml.fyi/taking-stock-23/
33 https://www.mosaicml.com/blog/mpt-7b

This has both:

1. Upsides: We know that public, high quality data sets can be created by volunteer communities, whether hosted by non-profits like Wikimedia and Archive or by for-profits like Reddit, Flickr, GitHub, and Stack Overflow. This gives volunteer, collaborative communities a standing that they lack in many other areas of tech policy; and
2. Downsides: Open data and open creative communities are totally unprepared for the trust and safety burden that has been thrust on them by their use in training. Compare how complex Wikipedia's trust and safety efforts are, compared to how non-existent LAION or C4's trust and safety efforts are. Similarly, many proprietary ML models are moderated by large, expensive, traumatised teams in places like Kenya."[34] 'Open' has no equivalent, or alternative, approach at this time, which may contribute to bias issues.[35]

Open AI is going to be regulated

> The biggest unknown for open AI is the nature of regulation. Because the work of understanding and influencing policy is specialised and expensive, open communities are at a significant disadvantage in any regulatory endeavour, as has been demonstrated recently in the UK, demonstrated recently in the UK by Signal's threat to leave the UK in response to the proposed "Online Safety Bill."[36]
>
> There are other challenges as well: ML regulation and judicial decisions will not be consistent from country to country. This is bad for open, which benefits from globally-sized communities. If communities need one open model for the US, another for Europe, another for China, etc., then many of the collaborative benefits of open will be lost.

ML regulation will move surprisingly quickly, with the EU already having voted on in-depth proposals on AI liability.

> Perhaps the biggest upside is open's transparency, which could give it a regulatory advantage: Transparency is a big emphasis of many ML regulatory proposals, particularly around training data and techniques. Open-native models and approaches are much better positioned to meet transparency requirements than closed models.

Open AI has a tonne of opportunities

The good news is that, unless regulation completely quashes it, Open AI has a tonne of opportunities at all levels. I expect we'll see vast amounts of investment in open AI, led by the model of London-based Stability.ai.[37] We'll also see innovation in the institutional space, with the distinct possibility of a European non-profit aggregating many data sources or models, to increase accountability—and reduce developer liability.

So interesting times are ahead!

34 Vice. (2023). OpenAI Used Kenyan Workers Making $2 an Hour to Filter Traumatic Content from ChatGPT. Retrieved from: https://www.vice.com/en/article/wxn3kw/openai-used-kenyan-workers-making-dollar2-an-hour-to-filter-traumatic-content-from-chatgpt
35 Bloomberg UK. (2023). Humans Are Biased.Generative Ai Is Even Worse. Retrieved from: https://www.bloomberg.com/graphics/2023-generative-ai-bias/
36 Guardian UK. (2023). Signal app warns it will quit UK if law weakens end-to-end encryption. Retrieved from: https://www.theguardian.com/technology/2023/feb/24/signal-app-warns-it-will-quit-uk-if-law-weakens-end-to-end-encryption
37 TechCrunch. (2023). France's Mistral AI blows in with a $113M seed round at a $260M valuation to take on OpenAI. Retrieved from: https://techcrunch.com/2023/06/13/frances-mistral-ai-blows-in-with-a-113m-seed-round-at-a-260m-valuation-to-take-on-openai/

3.4 Thought Leadership: What Open Data Means in the world of Generative AI

Sonia Cooper
Assistant General Counsel
Open Innovation
Microsoft

The importance of open data is back in focus. AI may well be the most consequential technology of our time, delivering advancements across all industries and sectors. The advent of foundation models for Generative AI, which require vast amounts of broad and diverse data, has renewed the importance of open data. Varied datasets enable AI models to learn patterns, recognise objects, make accurate predictions, and help people create. The larger and more representative the dataset, the better the AI model performance. Data, which can be broadly accessed, and which is made as open as possible, plays a pivotal role in ensuring that there is sufficient data available to train AI models, to facilitate innovation and ensure unbiased, performant AI.

Data accessbility and the widening divide

AI holds immense potential to revolutionise industries, enhance decision making, and address complex challenges. The opportunities for automation, optimisation, and prediction, empowering organisations and individuals to achieve greater efficiency and innovation are unprecedented. From healthcare and finance to transport and education, AI-driven solutions have the capability to improve outcomes, save costs, and enhance the overall quality of our lives. These opportunities will be unlocked only if organisations developing AI have sufficient access to broad and diverse data.

That data must be as accessible and usable as possible. There is, however, a widening divide between those who are able to put data to work and those who cannot. This data divide is due both to barriers to accessing data and to using data caused by a lack of data science and analytics expertise. The use of data tags to enable effective search functions is limited as found in a recent study.[38] Outside of the tech sector, fewer than 1% of companies have data scientists that can work with their data.

> For AI to benefit everyone they should be able to access and use the data they need to develop and use AI, so that research and development can occur across all sectors and for the benefit of all parts of society. Conversely, restricting access to data for the purpose of training AI will see research and development in AI restricted to furthering the priorities of a limited number of organisations that have access to data.

The opportunities presented by AI must be available to all if we are going to develop AI that benefits everyone. In this context, Generative AI presents particularly interesting opportunities and challenges. It presents us with the opportunity to democratise AI and the availability of data – by using large language models to develop conversational interfaces to AI functionality,

38 chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://arxiv.org/pdf/2306.06191.pdf

AI becomes easier for everyone to use. By using Generative AI to interact with data, the insights in data become more widely available to more people. Generative AI interfaces will make it easier to request feedback from open data users, thereby increasing the incentives to publish open data. As more data tools incorporate Generative AI, data users will be able to inform data publishers about how they are using the data, while they interact with it. As we find data gaps, Generative AI may even help plug those gaps by generating synthetic data or make it easier to publish data. Data publishing tools incorporating AI will also assist with the creation of metadata when publishing. This virtuous cycle underscores the exponential benefits when we are more open with data. Similarly, challenges presented from Generative AI arise from the greater dependency on vast amounts of data. Without access to enough data, AI will not perform well and this virtuous cycle cannot be created.

As we found in the open data steward peer learning network, with Microsoft and the Open Data Institute[39], feedback loops that show data publishers and other data users how data is being used encourage greater community engagement and more effective data publishing. This can include providing mechanisms for users to contribute back to the data sources, report errors, suggest improvements, or share derived insights. Such collaboration promotes data quality, ensures ongoing updates and maintenance, and strengthens the open data ecosystem. This is one of the principles explored in the Microsoft and the Open Data Institute's toolkit for new data publishers[40].

Open data practices and copyright

Issues surrounding data access are compounded by concerns among those who are not confident in being able to use data and in part due to questions as to whether copyright is infringed when copyrighted works are used to train AI. In 2022, the UK Government withdrew its decision to implement an explicit copyright exception for text and data mining, creating confusion regarding the ability to text and data mine in the UK. The UK Intellectual Property Office is now working on a code of practice for text and data mining to avoid the need for legislative change. This seems to be a good opportunity for the UK Government to clarify that exceptions already exist in UK legislation that permit text and data mining.

Performing text and data mining on works protected by copyright should not be considered a copyright infringement. As required under TRIPS, 9(2), copyright should not extend to ideas, procedures, or mathematical concepts. Everyone should have the right to extract knowledge from copyrighted works – to read, to learn, to understand, to develop ways to create new works and use technology as a tool to enable this.

Policy makers should take care not to attempt to reverse fundamental principles of copyright law which prevent copyright from extending to ideas and concepts. Some are suggesting that data made publicly-available on the internet cannot be used for text and data mining without a copyright licence, without regard to existing exceptions. This is a worrying development. Placing conditions on the use of data on the internet that extends beyond the limits of copyright risks further exacerbating the data divide, putting a cap on future innovations. It further attacks the notion of a free and open internet. We are only beginning to see the opportunities that AI presents. We must be careful not to limit our ability to share facts and ideas widely. Instead, governments should help foster a culture of openness and incentivise data sharing and usage, so we can unlock the full potential of AI and ensure that its benefits are accessible to all.

39 Open Data Institute. (2023). Reflections from our second peer-learning network with Microsoft. Retrieved from: https://www.theodi.org/article/reflections-from-our-second-peer-learning-network-with-microsoft
40 Open Data Institute. (2023). Supporting new data publishers to take a user-centric approach to data publishing. Retrieved from: https://www.theodi.org/project/supporting-new-data-publishers-to-take-a-user-centric-approach-to-data-publishing/

3.5 Thought Leadership: Building Better AI in the open

The Alan Turing Institute

Jennifer Ding
Senior Researcher
Turing Institute

Current approaches to openness in AI build upon years of development within the Open Source, open science, and open data movements by public and private institutions. Open principles have enabled alternative pathways to emerge for how AI is produced[41] compared to methods used in closed, for-profit AI labs. These approaches expand who can be part of the process and shape the values embedded in the technology. But how do we define the "open" in "open AI"?[42]

> Does open refer to un-gated access to the artefacts of machine learning, such as the data, trained weights, and model outputs? Or to the process of production - the code - and governance over decisions made by humans and machines? The latest AI advancements have obfuscated our understanding of how the technology functions and its impact on society, including what kinds of outcomes openness can catalyse.

As more projects and organisations begin to use the term "open AI" or "Open Source AI", the Turing's Tools, Practices and Systems (TPS)[43] research programme is convening our community to produce clearer definitions so that we know what to expect from AI systems badged as 'open'.

As part of the Open Source Initiative call for defining Open Source AI,[44] we are contributing a definition of openness that aligns with our approach to AI through safe, ethical, responsible, inclusive practices that follow the FAIR principles (Findable, Accessible, Interoperable, and Reusable) which we call "open AI."

Through AI collaboratives such as BigScience and EleutherAI, we have examples of openness beyond the sphere of major technology corporations in the US, which dominate AI narratives. While there are valid concerns about potential harms[45] that unrestricted openness can lead to, we'd like to highlight four beneficial outcomes that open AI can enable.

Open AI leads to faster innovation and better performance

Open AI models like Stable Diffusion and Falcon have been shown to match or beat performance of their closed counterparts like DALLE-2 and GPT-3, challenging long held beliefs that state-of-the-art models can only be produced by a handful of private technology companies

41 The Alan Turing Institute. (2023). Towards Openness Beyond Open Access: User Journeys through 3 Open AI Collaboratives. Retrieved from: https://arxiv.org/pdf/2301.08488.pdf
42 Ding, J. (2023). What defines the 'open' in 'open AI'?. The Alan Turing Institute. Retrieved from: https://www.turing.ac.uk/blog/what-defines-open-open-ai
43 The Alan Turing Institute. (2023). Tools, practices and systems. Retrieved from: https://www.turing.ac.uk/research/research-programmes/tools-practices-and-systems
44 OSI. (2023). Defining Open Source AI. Retrieved from: https://opensource.org/deepdive/
45 https://www.semianalysis.com/p/google-we-have-no-moat-and-neither

in even fewer countries. In a leaked memo, a member of Google staff has noted[46] how open AI developers, distributed around the world, are proving that smaller, fine-tuned models can outperform larger models, or in their words "eating our lunch".

We see that open AI is upending the misconception that "bigger is better" and that in order to build state-of-the-art one needs Internet-size datasets, deeper architectures, and ever longer training times. This phenomenon, embodied by the flurry of development sparked by the leaking of Meta's LLaMA[47] but also characteristic of past open releases for models like YOLO and BERT, demonstrates the pace and scale of innovation in the open AI community.

Open AI enables more equitable and empowering data practices

In 2021, Hugging Face co-organised the year-long Big Science Workshop[48], a collaboration of over 1,000 volunteer researchers from around the world to create the BLOOM[49] (short for BigScience Large Open-science Open-access Multilingual) Large Language Model (LLM). With its publicly documented progress, and its open invite to be part of the model licensing and data governance process, the BigScience team has distributed the power to shape, criticise and run an LLM to communities outside big tech. BigScience has also incorporated local needs through regional working groups that extend model localisation beyond just inclusion of a language to context-specific decision-making and evaluation. This transparent, collaborative approach to data governance may also be a way for AI collaboratives to avoid costly lawsuits, as seen through the examples of Stable Diffusion, Chat GPT and CoPilot.

Open Standards for AI build public trust

The confusion and hype surrounding AI releases may lead to building attention and usage for certain models, but do not support building public trust and understanding in the technology. Regulation like the EU AI Act[50] and initiatives like the Turing's AI Standards Hub[51], can counter this effect by contributing clear frameworks and accessible information about what characteristics of AI systems should be expected for use in different situations, particularly for safety-critical domains and public-facing decisions.

In recognition of the need for social scaffolding to enable widespread AI adoption, groups like Responsible AI Licences have been working on new forms of AI licences that balance openness and ethical concerns, driven by similar motivations as the ethical source community[52]. The Open-RAIL licence[53] has been adapted and adopted for open AI models like Stable Diffusion, BLOOM and StarCoder to address some of the challenges in licensing AI artefacts and maintaining the principles of openness for public transparency and access while addressing concerns about use for harmful purposes. Open-RAIL allows for the licensing of specific artefacts (e.g. Open RAIL-DS focuses on data and source code) and includes a use-based restrictions clause to specify what the prohibited uses include (e.g. illegal activity, generating misinformation). The RAIL community has worked with a range of organisations to evolve the licence to a trusted, usable form by many different kinds of actors to address concerns around flexibility, legibility, and liability.

46 NCBI. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792175/
47 The Verge. (2023). Meta's powerful AI language model has leaked online — what happens now? Retrieved from: https://www.theverge.com/2023/3/8/23629362/meta-ai-language-model-llama-leak-online-misuse
48 https://bigscience.huggingface.co/
49 https://huggingface.co/bigscience/bloom
50 The AI Act. (2023). The Artificial Intelligence Act. Retrieved from: https://artificialintelligenceact.eu/
51 https://aistandardshub.org/
52 https://ethicalsource.dev/
53 Responsible AI. (2022). From Rail To Open Rail: Topologies Of Rail Licenses. Retrieved from: https://www.licenses.ai/blog/2022/8/18/naming-convention-of-responsible-ai-licenses

Open AI enables more expansive and inclusive definitions of "better" AI

AI ethics experts have questioned[54] over the years whether models that are biassed against certain groups or require enormous expenditures of energy to train among other social harms should be considered "good enough" to deploy or make available for general use.

> Beyond the performance metrics the AI world uses to compare and rank models[55], open AI introduces new ways of deciding what "better" AI looks like and who is able to have a say. When more people – users, developers, and impacted and historically marginalised groups – can get involved in defining what characteristics make a model worthy of public attention and use, we will have better definitions and versions of AI in the world.

By challenging the status quo and expanding participation to invite artists, civil servants, philosophers, students, and more people into the conversation about and production of AI, open AI leads to more sustainable, more beneficial, and better performing models for the benefit of more people. Through open AI, we see a practice of openness that goes beyond expanding access[56], embodied through the work of projects like The Turing Way[57].

We see openness for faster innovation, for broadening participation, and as a response to addressing power imbalance, so that more people can be part of the creation process, and hence positively benefit from our AI futures.

54 https://dl.acm.org/doi/10.1145/3442188.3445922
55 https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard
56 https://zenodo.org/record/8028175
57 https://the-turing-way.netlify.app/index.html

3.6 Thought Leadership: A New National Purpose

TONY BLAIR INSTITUTE FOR GLOBAL CHANGE

Benedict Macon-Cooney
Chief Strategist
Tony Blair Institute for Global Change

We are in a new era. The progress in AI will almost certainly be the most important technological development of our lifetime. It will help with relatively simple things such as process automation while opening up profound possibilities for discovery in biology and material science, as well as for industries such as energy that are essential to our physical world.

As we set out in our recent paper, "A New National Purpose"[58], the pace at which AI Is developing presents both significant opportunities and risks for the UK. The country has some strength in depth in this technology, with companies such as DeepMind, Faculty and Exscientia contributing to accelerate nuclear fusion science, precision medicine and improve the machinery of government. Through the AI Taskforce and appointment of Ian Hogarth,[59] the UK government has also signalled its intent to bring deep expertise into decision-making.

But the scale of ambition needs to step up if the UK is to maintain an edge in the critical strategic technology of the 21st Century. The plan we put forward includes how to build AI-era infrastructure, improve talent, encourage innovation as well as how AI can change the shape and role of the state today. Our ideas included increasing compute capacity tenfold, new Disruptive Innovation Laboratories and an Advanced Procurement Agency so that the government can stimulate the UK ecosystem.

Given the pace of change and the uncertainty of the situation, we also believe it is too soon to lay out a definitive plan of action for regulation of AI. Instead, through a new national effort focused on researching and testing safe AI, we believe that research and regulation should proceed in tandem and in an agile way.

> If the UK is to be a global leader in AI, it is also important that the UK takes a clear position on the Open Source Community. The software that has been built by this community underpins the modern internet, from infrastructure to operating systems to algorithms. As a result of recent progress, Open Source is again enabling cutting-edge development in AI. Some of this will be driven by the reduction in compute costs, while the ability to run advanced models such as Stable Diffusion on a home computer puts powerful tools in the hands of artists, creators and small businesses. It will be integral to the democratisation of AI.

58 Tony Blair Institute for Global Change. (2023). A new national purpose: AI promises a world-leading future of Britain. Retrieved from: https://www.institute.global/insights/politics-and-governance/new-national-purpose-ai-promises-world-leading-future-of-britain
59 Gov.UK. (2023). Tech entrepreneur Ian Hogarth to lead UK's AI Foundation Model Taskforce. Retrieved from: https://www.gov.uk/government/news/tech-entrepreneur-ian-hogarth-to-lead-uks-ai-foundation-model-taskforce

Meanwhile, Meta's LLaMA allowed Stanford University to build a GPT-3-like model of their own for just $600. The Emirate of Abu Dhabi has also made its AI model, Falcon 40B, available Open Source for research and commercial use.

Such availability is important for students, researchers and builders around the world. "Open Source AI" speeds adoption, increases fair and trustworthy AI and advances the sciences that use AI too. But openness does also present risks, as it also provides bad actors with potential powerful tools.

> Some have taken pre-emptive steps to react to risks that are yet to unfold. The European Union, which tends to take a position of regulator of resort, has decided to weaken the position of the Open Source community through the AI Act. This would be a mistake and the UK should not follow suit. Instead, we should use the opportunity to offer a different model.

Providing cloud access to frontier API models would be one way of supporting Open Source Software in a way that mitigates the risk of misuse, and one we argued for in a New National Purpose. But there are other approaches too, such as the one pursued by the French National Centre for Scientific Research.

The Taskforce's role on this will also be critical.

And just as building in the open provides opportunities, building Open Source safety would confer significant benefits to the AI community and wider world.

The future of economic progress and growth is going to be dependent on AI.

> The UK has to show leadership in building its industry, including Open Source. This requires taking a clear position on the value that openness brings. There is a risk that this technology is confined to the hands of just a few actors, with the potential dividends being too narrow.

This has happened before with technology.

It cannot happen with the most transformative technology that exists today.

PART FOUR: CASE STUDIES

4.1 Case Study: Seldon

SELDON

Alex Housley
Founder & CEO

Seldon is a London-based AI company founded in 2018. It specialises in deploying and managing machine learning (ML) models in production, addressing the infrastructure challenges faced by organisations scaling their ML initiatives. Seldon’s core product, Seldon Core, has become an industry standard for model serving. They work with large enterprises, offering governance, risk compliance, and model lifecycle management solutions.

Seldon’s focus on ML model deployment and turning this into mass production has resonated with Fortune 500 company customers like Ford, PayPal, and Johnson & Johnson, as well as startups like Odyssey in the life sciences sector. Seldon’s aspiration is to bridge the gap between data scientists and DevOps teams, facilitating collaboration and reducing time-to-production for ML models.

While Seldon’s core technologies are Open Source, the company recognises that large enterprises often require additional assurances and support. To address this, Seldon offers fully packaged, warranted, and trusted software binaries. By providing commercial-grade products with service level agreements (SLAs) and warranties, Seldon offers a curation service to enterprises through a secure and reliable solution while benefiting from the stability and robustness of Open Source technologies.

Seldon’s focus is on a field called MLOps, which is about ensuring AI models work well in real-world situations. They help data scientists and technology teams collaborate more effectively and speed up the process of getting AI models into use without needing deep expertise in Kubernetes and DevOps. They also emphasise the importance of using high-quality data to improve the performance of AI models.

In the global AI conversation, Seldon recognizes the impact of Generative AI, taking the example of breakthroughs like ChatGPT. While Generative AI hasn’t caused a complete shift in adoption, it has opened up new possibilities and use cases. Seldon aims to empower a wider range of users, including engineers and developers, to leverage these advancements by providing user-friendly tools and integrating them into the software development process, improving workflows and productivity.

Seldon wants to make sure that people without technical expertise can benefit from Generative AI advancements. Combining Open Source technologies, enterprise-grade offerings and focusing on collaboration between data scientists and software developers, they have helped a lot of businesses allowing them to grow. They have built a strong customer base and strategic partnerships.

Overall, Seldon has been successful in helping businesses in the UK make the most of AI. Their focus on deploying AI models and their use of Open Source technologies have made them a trusted choice for companies. With their user-friendly tools and emphasis on collaboration and data quality, Seldon is making AI more accessible and useful for businesses in the UK of all sizes.

4.2 Case Study: Diff Blue

diffblue
AI for Code

Mathew Lodge
CEO

Born in the UK, Mathew spent the majority of his professional life in the Bay Area, where he worked with key players of the software industry, such as Cisco, Symantec, and VMWare, as well as small VC-backed start-ups, helping him gain invaluable experience in building products used by hundreds of thousands of organisations world-wide. All of this experience he put to good use when he moved back to the UK and became the CEO of Diffblue. Diffblue, head-quartered in Oxford, evolved from an Open Source project and turned into Generative AI for Code software.

A collaborative approach: Open Source and Diffblue

Diffblue spun out of the AI research group at the University of Oxford in 2016. Today With a team of 50 members, it has grown to include some of the world's largest banks, healthcare and tech companies as customers. Simply put, Diffblue is a Generative AI for code companies that produces software that writes software. As Mathew explains, demand for software developers far outstrips supply and the gap is set to get much wider over the coming decade. Today, near-ly 100% of code is written by hand, yet at least 50% of it is repetitive, error-prone and tedious to write. Diffblue's market-leading AI technology and approach can automate that 50%, un-locking enormous time and cost savings for its customers.

Mathew explains that Diffblue maintains a leading Open Source code analysis tool and sells a proprietary main product, and has a strong relationship with Open Source components. "Open Source is very important to our product in terms of the ingredients, and we contribute back to those projects. We're very interested in making sure that the language models that are used inside a product like ours are Open Source because we believe it's much healthier for the entire ecosystem. Our core product uses reinforcement learning for the most part and we use pieces of large language models in what we do."

Perks of the location

In terms of the benefits of being located in the UK, Mathew lists the affordability of talent. He compares it to the Bay Area and feels that the cost of creating a team there is much higher. The downside of the UK he feels is access to capital. He notes that there are far fewer venture firms that really understand the industry and the technology. As he goes on to say, "The UK is very good at very smart people and good ideas. The founders of Diffblue picked the right problem to solve, they made lots of good decisions about the problem and what to go after." Although he acknowledges that the single largest market for what they do is in the US - there's no getting around that. They plan on setting up in the US at some point next year.

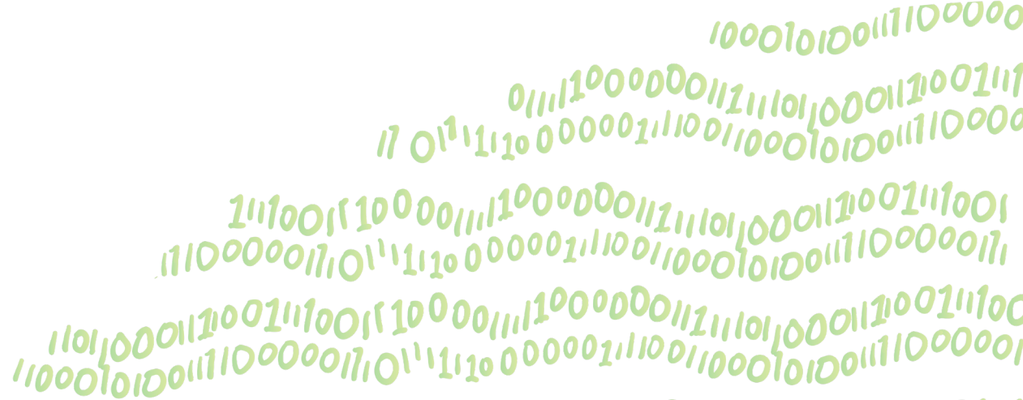
The explosion of AI

Diffblue is now seeing customers evaluating their product because of ChatGPT. There is now a huge consciousness of what Generative AI can do. It's been incredibly helpful to them because a lot of people didn't realise that they could buy software to solve a coding problem.

As he says, "They thought that hiring people was the only way. They didn't believe we could do what we said - they thought we were exaggerating. ChatGPT has really helped us with both of those problems, because now people know that it's possible - this software does exist, and it can do some incredible things. And this is where Open Source is really vital in AI."

Historically, foundational models were only accessible to people who had billions of dollars in cash, because it included very expensive training runs -- $4-$5m per run. The emergence of Open Source models that are pre-trained allows companies to adopt AI solutions in a wider way and fine tune those models for specific use cases.

As Mathew says, "I'm very excited about what could happen with foundational models with Open Source code and pre-trained weights that can be fine-tuned. Some of the innovations that we've seen in things like Open Source stable diffusion are a really great exam-ple. Where are the constraints - people don't have a lot of money, they don't have a lot of hardware. And so they've made stable diffusion -- the Open Source project -- much more efficient now. It's much easier and cheaper to train yourself, and you can run it your-self. There's a huge amount of innovation around that and it now out-performs systems made by multi-bil-lion dollar companies. And that to me, is a really great example of the power of Open Source."



4.3 Case Study: Aleois’ Quivr

Ben Ellerby
Founder
Aleois

Theodo Group invests in R&D with a focus on serverless, Generative AI through its Aleois company. Two of the three founders are in London. Aleois helps companies to adopt cloud strategies and its AI product is a web app called “Quivr”, which is an Open Source work in progress, less than two months old, with 50 plus contributors, and which is currently looking for a new name. It creates an all in one solution which “Stan Girard[60],” one of the group employees, dreamt up. The team are helping him to enact this vision, of giving anyone access to AI solutions. The whole modular back end is Open Source, almost 10,000 aApp is hosted at cost to Aleois and free to use to the end user.

Train it on your own body of knowledge, or you could give it a crm or spreadsheet, to allow the AI to train on it, then ask it for its suggestions. It will not upload sensitive documents, as there is a risk as with Gen AI, as to how secure data is. Focus on retrieving information with natural language, as if you are asking a friend. Next stage is to move to write documentation and be able to directly search it.

Building “separate brains” to ring-fence bodies of information, like sensitive and personal data. For now in the LLM space is one where companies are not ready to ensure that security which leaves the risk of data leakage which is something that companies need to react rapidly too, to ensure that their employees are not leaking data.

Quivr works off any LLM, as it is compatible with the top 5 LLMS - GPT 4 and 3.5, Vertex, Cord, and Open Source models Hugging Face and GPTforAll which was built off LLaMA. Its interoperability is its key advantage offering plug and play with limited customisation.

GitHub repo https://github.com/stangirard/quivr

60 https://www.linkedin.com/in/stanislas-girard/

“I am not a fan of attributing personality to systems. Even if we did come up with a genuine artificial intelligence that thought for itself there must be human responsibility. We’ve got a legislative precedent, of course, with computer generated art works, because the person who puts into place the systems for creating the artwork actually has the copyright. And I do think that at the copyright level that should be dealt with straightforwardly. You should start with an arbitrary rule that the person who creates the AI has the copyright in and is responsible for the output.”
Iain Mitchell, Honorary KC, OpenUK

4.4 Case Study: Progressio AI

PROGRESSIO
INSPIRE AI TRUST

Margaret Hartnett
Co-Founder

Margaret is an expert in deep tech with a PhD in the application of AI in analytical chemistry. She wears many hats, including an award-winning technologist, patent attorney and business leader who bridges the gap between commercial, legal and technology domains. After spending over a decade in academia, she decided to leap head-first into the .com boom and became a patent attorney and IP specialist in the software industry. During this time, Margaret maintained a very close working relationship with AI, as many of the people she represented were AI providers. Her close contact with the AI tech scene prompted Margaret to leave private practice and join an AI tech scale-up in London where she started off as their patent counsel and quickly became their director of legal and operations.

Spotting a gap in the market, Margaret combined her legal and technical knowledge to co-found Progressio AI, eight months ago. Progressio AI is now on its way to become a leading AI Governance technology provider, that offers tailored solutions to enable businesses to navigate the complexities of AI-related regulation and legislation.

Innovation, patents and Open Source

In Margaret's professional experience of the software industry, "if somebody tells you that they're not using Open Source, they're either lying, deluded, or they just don't know". According to her, in reality, everyone in the industry uses Open Source Software, either in their products or as supporting tools at various stages during the development of their products - resulting in innovative outputs and results. She believes that freedom of choice remains with the creator, "if they want to make their software available as Open Source, wonderful, if they want to patent it, similarly wonderful. But you can't talk about developing software without talking about Open Source. Open Source is just as important to any aspect of the discussion of intellectual property as copyright, trademarks or patents."

Ticking all the boxesL the adoption of AI

Recognising the obligations that the European AI Act puts on AI providers, Margaret and her co-founders established Progressio AI to fill the potential need it will create, saying, "the AI Act essentially requires AI providers and AI deployers to document every stage in the lifecycle of an AI system. And, from my experience, both as an AI researcher and as a person who has worked with and in the AI industry for decades, I strongly believe that most AI companies are unprepared for compliance with the AI Act ."

The emergence of ChatGPT is a quantum leap forward in the evolution of AI. However, the slower pace of legislation and regulation creation is leading to a disconnect between the adoption of AI technologies and their governance. The AI Act "imposes very significant obligations on providers along the value chain of large language models (LLMs)" bringing into question the ownership of their outputs. A critical area, for example, is attribution of software created by AI trained on Open Source works. The norm is for Open Source licences to require attribution. She explains, "to my mind, it doesn't matter whether it's an LLM or a human being, if they are using an Open Source work, the contribution that the author of the work has very kindly made should be respected. They didn't have to make their work available as Open Source. Authors should be recognised for the time, effort and creativity they put into their works. Following the requirements in Open Source licences, recognises people's creativity."

But how do you determine the extent of its use? Can you really say that you can ascribe Open Source licence terms from works that might be 200 lines of code, when the LLM might only be using one part of one line of code from the work. That is really difficult.

As she says, "It's a challenging one, because so many of the really great models that we're seeing today are coming from the Open Source community. These evolutions in AI wouldn't exist without Open Source, and certainly not without the open economy, open data, open publications, etc. The distinction though, between a model and its application in business is immense.

That's one of the issues we deal with in Progressio AI " our aim is to support organisations answering such questions, thereby ensuring that AI is developed, deployed and leveraged in a responsible and ethical way".

Progressio AI is developing tools to automate various aspects of compliance with the AI act, to make it easier for AI companies to comply. Because, as she says, "AI governance is not a one-off activity. It's not like you just get an audit done and that's it. On the contrary, AI governance is a process that continues during the entire lifecycle of an AI system."

4.5 Case Study: Evidently AI

EVIDENTLY AI

Elena Samuylova
Co-Founder

Elena is CEO and co-founder of Evidently AI, a company developing Open Source Software to help monitor machine learning models in production. Her background includes working at the Russian search engine, Yandex and five years focused on applying machine learning in different industries, from telecom and retail to more traditional internet applications. Overall she's spent 10 years on the applied side of machine learning. Both she and her co-founder are based in London with a dispersed team working from Argentina to the Netherlands. The organisation doesn't have office space.

Evidently's MLOps tool

Users of Evidently's Open Source Software product come from a variety of companies - both startups and enterprises, a lot of ecommerce, FinTech, banking, retail, even manufacturing, so it's very broad. The company doesn't develop AI but helps others who work with AI and machine learning. Companies run different types of models for example, demand forecasting models used to optimise logistics, or a marketing personalisation model and maybe a text classification model. All of these power the business' back end, some of them are user facing, some of them internal. When these models are in production, Evidently's tool helps ensure that they actually work and deliver the value they expect. This is part of what is known as MLOps - machine learning operations. It's a technology stack on the back end of all things that are happening throughout the ML model lifecycle, from model creation to model deployment. Evidently exists in after Model Deployment and is a Python library that is used to enable visibility into production model performance. .

AI and Open Source

> AI and Machine Learning technologies sit on top of Open Source Software by design. Most tooling used to develop these systems is Open Source. It's been this way for the last decade - it's not something new happening now. It's almost expected by any developer, data scientist, machine learning engineer that works with this technology, that most of the tools will be available Open Source. Typically, you expect your data to be proprietary, but in creating these models, you use Open Source Software.

The new part that's happened is that we are talking about large pre-trained models that can be also in Open Source. This is the new development and huge discussions relate to the LLMs, Chat-GPT and everything that's happening today.

Elena's strong belief is that for AI and machine learning it's impossible to create AI without Open Source. In the last six months Chat-GPT 4 created a very easy interface to interact with these sorts of systems which popularised the idea.
She notes, "There was no new technological breakthrough, it's more that this technology has become known and available to so many companies and people that are now experimenting and thinking how to develop it. It's more about a marketing spike that has happened, providing the tool in a very accessible form."

> She sees automation as having the biggest impact - automating some specific decisions or processes powered by data and machine learning to help you achieve a lot of different operational efficiencies. These operational efficiencies and productivity are the key to the benefits being created. Chat-GPT and other LLMs bring actual use cases across different industries - that's the most exciting part, the application.

> She feels very strongly about disclosure of datasets, not just from the legal standpoint but because disclosure gives a lot of insight to potential users on model capabilities and biases. This is very important for pragmatic use of the model to be successful for anyone. Understanding what the model might know means understanding what type of data sources and curation process it has experienced. There is a lot of interesting work around ML model cards, Data Set cards and data sheets. These are basically documents that describe what the data the model was trained on - even when talking about large language models.

It would be interesting to talk not just about Open Source AI, but about Open Source data sets and Open Source Models, because these are the two most interesting things in this space and they might exist separately.

PART FIVE: CONCLUSION

5.1 Show the UK the Money

Dr Jennifer Barth
Chief Research Officer
OpenUK

The conversation around AI is only escalating. It feels disembodied, objective and inevitable. But in fact, the escalation of the conversation, the constantly morphing narrative, the fast paced innovation in the form of successive versions are all indicative of AI’s materiality and a reflection of human ideologies. We are, it seems, in the middle of a major innovative moment in AI, fuelled by simultaneous advances in data, models and machine learning. But if we are to have any opportunity of harnessing the power and potential of these advances in a meaningful way for humanity, we need to develop and deploy these systems in a way that aligns with human values and goals. As Margaret Hartnett notes in her case study above, “The distinction between a model and its application in business [and, I would add in everyday life] is immense.”

Questions of copyright, ownership, and the provenance of data, for example, are all questions that allow us to consider AI’s interaction with humans. Asking in our survey if Open Source offers the solution to concerns about AI ownership, 40% of the respondents agreed but, and more importantly, 45% are on the fence. As the report discusses, there is much debate as to the relationship between Open Source and AI and how we are to define this relationship is key to the next steps.

Much of the discussion of that relationship centres around Open Source Software, open data and open innovation, in terms of its values - transparency, trust, creativity, and, importantly, collaboration - all come to the fore in the meeting of Open Source and AI. Opening the software, data and potentially the innovations offer two important if not critical benefits to AI - transparency and the potential to democratise its development. 85% of respondents in our survey believe that datasets used to train an AI should be disclosed in the software supply chain - the hope for transparency to spot bias, inform copyright and origins is definitely on the agenda.

Can Open Source democratise AI and remove the risk of an arms race to AI solutions amongst a few big companies? This depends on finding alignment with human (and Open Source) values. The UK government’s most recent policy paper calls for AI Regulation: a pro innovation approach – setting out the UK government approach to “implementing a proportionate, future-proof, and pro-innovation framework for regulating AI.” They have also pledged to become a global leader in AI. The way forward echoes what Jennifer Ting says above, it’s about safe, ethical and responsible use and the UK could lead the way on this in its approach to regulation and governance, but more than that, on the excellent businesses working in AI focusing their build and deploy methods through and with a responsible lens.

At Symmetry, after years of experience implementing these models, we arrived at a framework for alignment with human endeavour - that is a set of responsibility domains, a lens through which it becomes possible to create intriguing experiments and investment of resources into ways of working to transform impact on all of these domains. These include responsibility to

the users, to the business, to the workforce, to the environment and to governance. For example,the often reduced compute requirements of Open Source models can greatly reduce the carbon footprint and ESG accounting of organisations pursuing clean and green engineering in their digital code bases and infrastructure. Likewise, greater transparency, accountability, trust and information security can come from organisations deploying Open Source solutions without some of the black box and opaque terms and conditions of proprietary software.

Open Source is asserting itself, galvanised and catalysed by the opportunity open innovation affords on the frontiers of an emerging technology domain like Generative AI.

PART SIX: FORMALITIES

6.1 Contributors

Alex Housley, Founder and CEO, Seldon
Alex is a deep tech entrepreneur who founded Seldon in 2014, after experiencing first-hand the challenges of scaling model deployment in leading companies. He is also a guest advisor at the All-Party Parliamentary Group, which explores the impact and implications of artificial intelligence and is at the forefront of informing policy-making in the UK.

Amanda Brock, CEO, OpenUK
Amanda is CEO of OpenUK and Executive Producer of State of Open Con https://stateofopen-con.com Amanda is a Board Member of the Open Source Initiative; appointed member of the Cabinet Office’s Open Standards Board; Member of the British Computer Society Inaugural Influence Board; and European Representative of the Open Invention Network. A lawyer of 25 years’ experience, she previously chaired the Open Source and IP Advisory Group of the United Nations Technology Innovation Labs, sat on the OASIS Open Projects and UK Government Energy Sector Digitalisation Task Force Advisory Boards. She was General Counsel of Canonical for 5 years from 2008 and set up their legal function. Amanda is a judge in the IDG Foundry CIO 100, 2023, having been a Judge in the “We are Tech Women Rising Star Awards” 2020-22. Awarded a Lifetime Achievement Award in Women, Influence & Power UK 2022,included in Computing’s 100 IT Leaders of 2023, Computer Weekly’s Most Influential Women in Tech Long list in 2021 and 2022 and in their UKTech50 Influencers longlist 2021-2023, and in the Involve HERoes list of 100 global Women executives driving change by example. She is the editor of Open Source Law, Policy and Practice (2nd edition) published by Oxford University Press in October 2022, with open access sponsored by the Vietsch Foundation.

Ben Ellerby, Founder, Aleois
Ben Ellerby is the Founder of Aleois and a dedicated member of the developer community. In 2020, AWS named him a Serverless Hero for his community and Open Source work. He is the editor of Serverless Transformation publication. Ben co-organizes the Serverless User Group in London, is part of the ServerlessDays organising team, and regularly speaks at technology conferences around the world. At Aleios and Theodo Group, he helps startups disrupt and assists large organisations in remaining competitive by leveraging cutting edge technology to solve real world problems. He advises several open-source Generative AI projects, as well as companies and NGO.

Benedict Macon-Cooney, Chief Policy Strategist, Tony Blair Institute for Global Change
Benedict is the Chief Policy Strategist at the Tony Blair Institute. His work focuses on how to harness the potential of technology to reform government and stimulate industries such as AI, biotech and clean tech. He has a background in economics, starting his career in HM Treasury, but has also worked in leaders offices around the world, including the President’s Office in Rwanda.

Christopher Eastham, Partner Field Fisher and OpenUK Chief Legal Officer
Chris is a Partner in the Technology & Data group at the law firm Fieldfisher. In his legal practice, he advises both customers and vendors on technology-related matters, providing commercial and pragmatic legal solutions to enable them to achieve their strategic goals. He is an expert in negotiating complex technology contracts—especially for business critical systems and functions. He has in-depth experience in the telecoms sector, and of AI, automation, robotics, AR/VR/XR, and other emerging technologies. Chris has first class degrees in Physics with Space Science and Technology (MPhys), and in Law (LLB).

Elena Samuylova, CEO and Co-Founder, Evidently AI
Elena is a CEO and Co-founder at Evidently AI, a Y Combinator-backed startup developing open-source tools to evaluate, test, and monitor the performance of machine learning models.

She has been active in the applied machine learning space since 2014. Previously, she co-founded and served as a CPO of an industrial AI startup. She worked with global metal and chemical companies to implement machine learning for production optimization. Prior to that, she led business development at Yandex Data Factory, an enterprise AI division of Yandex. She focused on delivering ML-based solutions to retail, banking, telecom, and other industries. In 2018, Elena was named 50 Women in Product Europe by the Product Management Festival.

Iain G Mitchell, KC, Tanfield Chambers and OpenUK Honorary KC
Iain graduated from Edinburgh University in 1973 and was called to the Scottish Bar in 1976, taking silk in 1992. He was called to the English Bar in 2012. He is recommended in the Legal 500 and Chambers Directory for IT Law, IP Law, and commercial dispute resolution.As well as being the Chairman of the Scottish Society of Computers & Law, he is also Chair of the CCBE Surveillance Working Group, a Member of the CCBE IT Committee; a member of the Bar Council IT Panel, a Freeman of the Worshipful Company of Information Technologists and serves on the Electronic Evidence Working Party of the Centre for European Policy Studies. He holds an honorary lectureship in IT Law in the ITM Institute at the Wilhelms-Universität, Münster and his publications include chapters in Electronic Evidence (Butterworths) and in Open Source Software: Law, Policy and Practice (OUP). He is also one of the joint editors of JOLTS.

Dr Jennifer Barth, Founder and Research Director, Symmetry
Jenn has more than 15 years of experience leading independent research on the intersections of emerging technologies and socioeconomic change. She provides companies with independent thought leadership and media engagement opportunities on global issues impacting and shaping our current and future technical-social lives. Her work spans the digital through to social and economic change. She has looked at sustainability, workforce skills and organisational competitiveness strategies through and beyond the pandemic with Microsoft and many other big and small organisations and works as the Chief Research Office researching the role of Open Source Software and its potential to fuel the circular economy with OpenUK. She has experience working on the human impact of artificial intelligence (AI) through fieldwork experiments with IBM Watson, Microsoft and other providers. She is skilled at blending research methods and working with people to bring to life the stories behind numbers. Dr Barth earned her DPhil in Geography from the University of Oxford.

Luis Villa, Co-Founder, Tidelift; Author, OpenML.fyi newsletter
Luis has been involved in Open Source since the late 90s, first as a developer and then as an attorney and community leader. He’s advised clients including Mozilla, the Wikimedia Foundation, tiny startups, and the largest Silicon Valley titans. He’s also served open communities including Creative Commons, the Open Source Initiative, the Open Knowledge Foundation, and Open Street Maps. He is currently the co-founder of Tidelift, a startup working to make Open Source better for maintainers and enterprises, and writes the OpenML.fyi newsletter, documenting the intersection of open software, open data, and machine learning.

Margaret Hartnett, Co-Founder, Progressio AI
Margaret holds a PhD in AI & Analytical Chemistry, Dr Margaret Hartnett is a senior researcher, European Patent Attorney, Chartered Patent Attorney and company director. With a track record of growing and exiting a hugely successful AI scale-up, Margaret has spent almost 20 years developing into an executive-level research, innovation and Intellectual Property (IP) specialist.

Bringing a rare interdisciplinary skill and perspective, Margaret bridges the gap between the commercial and technology functions of fast-growth, high-tech businesses. As co-founder of Progressio AI, she brings this expertise to bear to help AI-driven companies meet their obligations under the forthcoming EU AI Act.

Matt Barker, Global Head of Cloud Native Services, Venafi / Co-founder Jetstack and OpenUK Entrepreneur in Residence
Matt is passionate about how Open Source can be used as a driving force for value and innovation. Having spent his entire career working for Open Source start-ups, including Canonical and MongoDB he went on to start one of his own, Jetstack.Jetstack began as a bootstrapped Kubernetes service provider, and helped all manner of companies from small start-ups to large enterprises. In the process of spotting gaps around Kubernetes, they created the Open Source cert-manager project which has thousands of Github stars and is downloaded more than a million times a day. Jetstack was acquired by Venafi in 2020, the leader in machine identity management. Matt now runs their cloud native services business and continues to explore cutting edge technologies in the Open Source ecosystem.

Matthew Lodge, CEO, DiffBlue, Generative AI for Code
Mathew has been building software and hardware products for over 25 years and has done everything from writing code that flew on the Space Station to connecting 6 countries to the Internet for the first time, to building and launching VMware’s first cloud computing services. Before joining Diffblue as CEO, he spent over 20 years in Silicon Valley building products for Cisco, Symantec, VMware and VC-backed start-ups.

Sonia Cooper, Assistant General Counsel, Open Innovation Team, Microsoft
Sonia has over 25 years of experience advising on IP and innovation in the tech sector, specialising in AI and IP. A Chartered and European Patent Attorney, Sonia worked extensively with Microsoft Research in Cambridge and now leads IP policy in the Open Innovation Team at Microsoft. Sonia is Vice President and Immediate Past President of IP Federation and Chairs the federation’s Data and Copyright Committee. Sonia is the Data Policy lead for Open UK and is a passionate advocate for open data.

6.2 About the Creators of this Report

6.2.1 OpenUK

OpenUK is the organisation for the business of Open Technology in the UK, being Open Source Software, Open Source hardware and open data. As an industry organisation, OpenUK gives its participants greater influence than they could ever achieve alone. Open UK’s purpose is to promote UK leadership and global collaboration in Open Technology.

OpenUK is committed to promoting UK leadership in Open Technology and supporting collaboration between businesses, public sector organisations, government and communities to expand the opportunities available to all around Open Technology on a global basis. OpenUK creates a visible Open Technology community in the UK, and uses that community’s impact to ensure that the UK’s laws and policies work for Open Technology whilst encouraging the future community in the business of Open Technology through learning. OpenUK is a not-for-profit company limited by guarantee, company number 11209475 with its registered office at 8 Coldbath Square, London EC1N 5HL, www.openuk.uk, contact admin@openuk.uk

6.2.2 Symmetry

Symmetry looks beyond the surface and behind the curtain of the fundamental innovations and trends shaping our society, markets, culture, and values. We are academics and researchers looking at the intersections of emerging technology and socioeconomic impact, producing independent research for thought leadership and business solutions.

Symmetry’s mission is to share and grow knowledge about everyday lives. We want to understand the past, present, and future of human interaction with emerging technologies and socioeconomic changes—from behaviour to context, nature to nurture, origin to experiences—helping our clients engage their clients and public imagination.

6.2.3 Runa Capital

Runa Capital is a Luxembourg-based global venture capital firm, focusing on enterprise software, deep tech and fintech startups. Since 2010, we have invested in more than 100 European and US startups, including over ten Open Source companies.

Runa’s early-stage investments include cloud banking platform Mambu ($5.5B last round valuation), Open Source web server Nginx (acquired by F5 for $700 million), cloud ERP vendor Acumatica (acquired by EQT) and quantum computing company Pasqal.

Runa has supported UK startups like Zopa, DigitalGenius, Chattermill, Evidently AI and Lumai. Its London-based general partner Konstantin Vinogradov focuses on AI and Open Source.

6.3 Methodology

The research used a mixed method approach to explore and demonstrate the state of the Open Source Software economy - its value and values - in the UK. Interviews were conducted with industry leaders, founders and Open Source Software experts and included as case studies and thought leadership on the value of Open Source Software.

The research used a proximity sampling for the third annual OpenUK survey, receiving a total of 339 answers. Response collection took place in June 2023. Out of the 339 answers, sorting for those coming from UK based organisations (the focus of the research), there were 244 by participants from across the UK, representing all sectors of the economy. All percentages reported have been rounded, and where replies were fewer than 3 we have aggregated replies where possible to control for disclosure. Responses came from all over the UK, and those from abroad (16%) were filtered out.

Where are you based?

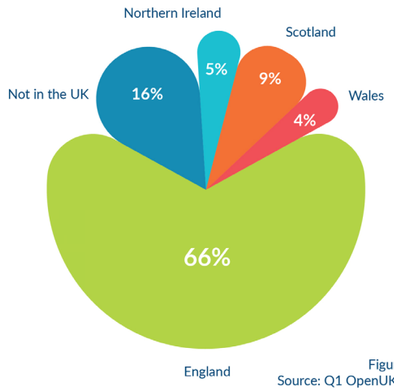


Figure 25. Where are you based?
Source: Q1 OpenUK State of Open Survey 2023

Out of those, the headquarters of the organisations of more than ⅔ (69%) are in the UK, while 31% reported that their organisation has headquarters overseas.

Location of HQ

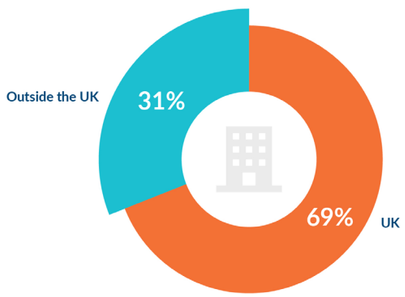


Figure 26. Location of HQ
Source: Q5 OpenUK State of Open Survey 2023

For some of the questions we received a lower number of answers - when excluding "I don't know" options, and "none" answers (especially in Q14). These were Q9 - 193 valid replies, Q13a - 201 valid replies, Q14 - 186 valid replies, Q15 - 156 valid replies. Once we applied a filter removing "none" responses in Q14, we obtained 103 responses for which we had all the necessary variables in a valid form. These are essential questions for the estimation of investment in Open Source Software. Consequently, due to the lower response rate which produced an even smaller sample size, there is a relatively lower level of confidence in the data.

The methodology of the investment estimation is outlined in OpenUK State of the Open: the UK in 2022. To calculate sectoral GVA, we used responses from the industrial categories "Information and Communication" and "Professional, Scientific and technical activities", to approximate the sector "IT, software and computer services industry". Then using estimates from the Department for Digital, Culture, Media and Sport (DCMS) for the 2022 GVA of the UK (£50.71 billion) and the average proportion of businesses revenue from Open Source Software in our survey (approximately 27%), we calculated GVA from Open Source Software. Confidence on this estimate is medium-low and should be interpreted with caution, as although the two industrial categories selected from the sample give an adequate number of businesses, the survey sampling method (proximity sampling) is not representative of the UK economy.

6.4 Acknowledgements

The research was led by Dr Jennifer Barth, CEO and Research Director at Symmetry and OpenUK's Chief Research Officer in partnership with Amanda Brock, CEO OpenUK in 2023. Thank you to our team of economists, psychologists, data scientists and social scientists to all who contributed, and in particular Zin Nwe Zaw Lwin, Dr Dora Kokosi and Matthew Buck of Drawnalism.

We are grateful to the 339 individuals who participated and provided us with essential data in our survey.

6.5 References

- Ada Lovelace Institute and the Alan Turing Institute. (2023). How do people feel about AI? Retrieved from: https://www.adalovelaceinstitute.org/report/public-attitudes-ai/
- Andreessen, M. (2023). Why AI Will Save the World. Retrieved from: https://a16z.com/2023/06/06/ai-will-save-the-world/
- Bloomberg UK. (2023). Humans Are Biased.Generative Ai Is Even Worse. Retrieved from: https://www.bloomberg.com/graphics/2023-generative-ai-bias/
- chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://arxiv.org/pdf/2306.06191.pdf
- Civo. (2022). The Kubernetes State of Play 2022. Retrieved from: https://www.civo.com/kubernetes-state-of-play-2022
- Ding, J. (2023). What defines the ‘open’ in ‘open AI’?. The Alan Turing Institute. Retrieved from: https://www.turing.ac.uk/blog/what-defines-open-open-ai
- European Commission. (2022). Cyber Resilience Act. Retrieved from: https://digital-strategy.ec.europa.eu/en/library/cyber-resilience-act
- Github Blog. (2023). The economic impact of the AI-powered developer lifecycle and lessons from GitHub Copilot. Retrieved from: https://github.blog/2023-06-27-the-economic-impact-of-the-ai-powered-developer-lifecycle-and-lessons-fromgithub-copilot/
- Github. (2023). Open Data on GitHub: Unlocking the Potential of AI. Retrieved from: https://arxiv.org/pdf/2306.06191.pdf
- Gov.UK. (2023). Tech entrepreneur Ian Hogarth to lead UK’s AI Foundation Model Taskforce. Retrieved from: https://www.gov.uk/government/news/tech-entrepreneur-ian-hogarth-to-lead-uks-ai-foundation-model-taskforce
- Gov.UK. (2023). UK to host first global summit on Artificial Intelligence. Retrieved from: https://www.gov.uk/government/news/uk-to-host-first-global-summit-on-artificial-intelligence
- GovUK. (2023). Tax relief for investors using venture capital schemes. Retrieved from: https://www.gov.uk/guidance/venture-capital-schemes-tax-relief-for-investors
- Guardian UK. (2023).
- Signal app warns it will quit UK if law weakens end-to-end encryption. Retrieved from: https://www.theguardian.com/technology/2023/feb/24/signal-app-warns-it-will-quit-uk-if-law-weakens-end-to-end-encryption
- https://aistandardshub.org/
- https://cert-manager.io/
- https://dl.acm.org/doi/10.1145/3442188.3445922
- https://ethicalsource.dev/
- https://huggingface.co/bigscience/bloom
- https://the-turing-way.netlify.app/index.html
- https://www.gnu.org/licenses/copyleft.en.html
- https://www.itworldcanada.com/post/significant-gravitas-launches-auto-gpt
- https://www.mosaicml.com/blog/mpt-7b
- https://www.openml.fyi/
- Mayo, M. (2023). The Role of Open Source Tools in Accelerating Data Science Progress. Retrieved from: https://www.kdnuggets.com/2023/05/role-open-source-tools-accelerating-data-science-progress.html
- MetaAI. (2023). Introducing LLaMA: A foundational, 65-billion-parameter large language model. Retrieved from: https://ai.facebook.com/blog/large-language-model-llama-meta-ai/
- NCBI. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792175/
- OECD. (2016). Measuring Investment In ICT Services Industries. Retrieved from: https://one.oecd.org/document/DSTI/CDEP/MADE(2016)5/en/pdf
- OECD.AI (2023). Visualisations powered by JSI using data from Github, accessed on 27/6/2023. Retrieved from: www.oecd.ai . Supported by the Patrick J. McGovern Foundation.
- ONS. (2023). Artificial Intelligence (AI) awareness, use and impact, Great Britain. UK Office for National Statistics. Retrieved from: https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/datasets/artificialintelligenceaiawarenessuseandimpactgreatbritain
- ONS. (2023). GDP quarterly national accounts, UK. Retrieved from: https://www.ons.gov.uk/economy/grossdomesticproductgdp/bulletins/quarterlynationalaccounts/januarytomarch2023
- ONS. (2023). Public opinions and social trends, Great Britain: 4 to 14 May 2023. Office for National Statistics. Retrieved from: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/bulletins/publicopinionsandsocialtrendsgreatbritain/4to14may2023
- Open Data Institute. (2023). Reflections from our second peer-learning network with Microsoft. Retrieved from: https://www.theodi.org/article/reflections-from-our-second-peer-learning-network-with-microsoft

- Open Data Institute. (2023). Supporting new data publishers to take a user-centric approach to data publishing. Retrieved from: https://www.theodi.org/project/supporting-new-data-publishers-to-take-a-user-centric-approach-to-data-publishing/
- Open Knowledge Foundation. (2023). Debriefing the Open Definition workshop at RightsCon. Retrieved from: https://blog.okfn.org/2023/06/26/debriefing-the-open-definition-workshop-at-rightscon/
- Open Logic by Perforce an Open Source Initiative. (2023). State of Open Source Report. Retrieved from: https://www.openlogic.com/resources/2023-state-open-source-report
- Open Source Initiative. (2023). Deep Dive: AI. Retrieved from: https://deepdive.opensource.org/wp-content/uploads/2023/02/Deep-Dive-AI-final-report.pdf
- OpenML. (2023). Taking stock of open(ish) machine learning / 2023-06-15. Retrieved from: https://www.openml.fyi/taking-stock-23/
- OpenUK. (2022). State of Open: The UK in 2022 Phase One. July 2022 https://openuk.uk/wp-content/uploads/2022/10/OpenUK-report_phase-03_FINALMASTER_V6.pdf
- OpenUK. (2022). State of Open: The UK in 2022, Phase 1. Retrieved from: https://openuk.uk/stateofopen/
- OpenUK. (2022). State of Open: The UK in 2023, “A Year in Review.” Retrieved from: https://openuk.uk/stateofopen/
- OSCI. (2020). Evaluating Open Source Trends from 2020 to Today. Retrieved from: https://opensourceindex.io/p/evaluating-open-source-trends-from-2020-to-today
- OSI Blog. (2023). Now is the time to define Open Source AI. Retrieved from: https://blog.opensource.org/now-is-the-time-to-define-open-source-ai/
- OSI. (2023). Defining Open Source AI. Retrieved from: https://opensource.org/deepdive/
- Responsible AI. (2022). From Rail To Open Rail: Topologies Of Rail Licenses. Retrieved from: https://www.licenses.ai/blog/2022/8/18/naming-convention-of-responsible-ai-licenses
- Reuters. (2022). U.S. Energy Department to spend $3.7 billion on carbon removal. Retrieved from: https://www.reuters.com/business/sustainable-business/us-energy-department-spend-37-billion-carbon-removal-2022-12-13/
- Runa Capital. (2023). The Fastest Growing Open Source startups in 2023. Retrieved from: https://runacap.com/ross-index/q1-2023/
- Semianalysis. (2023). Google “We Have No Moat, And Neither Does OpenAI.” Retrieved from: https://www.semianalysis.com/p/google-we-have-no-moat-and-neither
- Tech Crunch. (2022). Decentralized discourse: How Open Source is shaping Twitter’s future. Retrieved from-https://techcrunch.com/2022/12/13/decentralized-discourse-how-open-source-is-shaping-twitters-future
- TechCrunch. (2023). ChatGPT: Everything you need to know about the AI-powered chatbot. Retrieved from: https://techcrunch.com/2023/07/10/chatgpt-everything-you-need-to-know-about-the-open-ai-powered-chatbot/
- TechCrunch. (2023). France’s Mistral AI blows in with a $113M seed round at a $260M valuation to take on OpenAI. Retrieved from: https://techcrunch.com/2023/06/13/frances-mistral-ai-blows-in-with-a-113m-seed-round-at-a-260m-valuation-to-take-on-openai/
- The AI Act. (2023). The Artificial Intelligence Act. Retrieved from: https://artificialintelligenceact.eu/
- The Alan Turing Institute. (2023). Tools, practices and systems. Retrieved from: https://www.turing.ac.uk/research/research-programmes/tools-practices-and-systems
- The Alan Turing Institute. (2023). Towards Openness Beyond Open Access: User Journeys through 3 Open AI Collaboratives. Retrieved from: https://arxiv.org/pdf/2301.08488.pdf
- The Verge. (2023). Meta’s powerful AI language model has leaked online — what happens now? Retrieved from:
- Tony Blair Institute for Global Change. (2023). A new national purpose: AI promises a world-leading future of Britain. Retrieved from: https://www.institute.global/insights/politics-and-governance/new-national-purpose-ai-promises-world-leading-future-of-britain
- UKTN. (2023). Jeremy Hunt says he wants to make the UK the next Silicon Valley (again). Retrieved from: https://www.uktech.news/news/government-and-policy/hunt-uk-silicon-valley-20230127#:~:text=Hunt%20previously%20said%20he%20wants,the%20world’s%20next%20Silicon%20Valley%E2%80%9D.
- Vice. (2023). OpenAI Used Kenyan Workers Making $2 an Hour to Filter Traumatic Content from ChatGPT. Retrieved from: https://www.vice.com/en/article/wxn3kw/openai-used-kenyan-workers-making-dollar2-an-hour-to-filter-traumatic-content-from-chatgpt
- VMware. (2022). The State of the Software Supply Chain: Open Source Edition 2022. Retrieved from: https://tanzu.vmware.com/content/ebooks/state-of-software-supply-chain-2022
- WH Government. (2021). Executive Order on Improving the Nation’s Cybersecurity. Retrieved from: https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/
- Wikipedia. (2023). Reinforcement learning from human feedback. Retrieved from: https://en.wikipedia.org/wiki/Reinforcement_learning_from_human_feedback

6.6 Who’s Quoting State of Open?

We have been delighted to see our work attributed by:

6.6.1 McKinsey Supply Chain September 2022

https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/cybersecurity/software-bill-of-materials-managing-software-cybersecurity-risks

Why an SBOM program is necessary

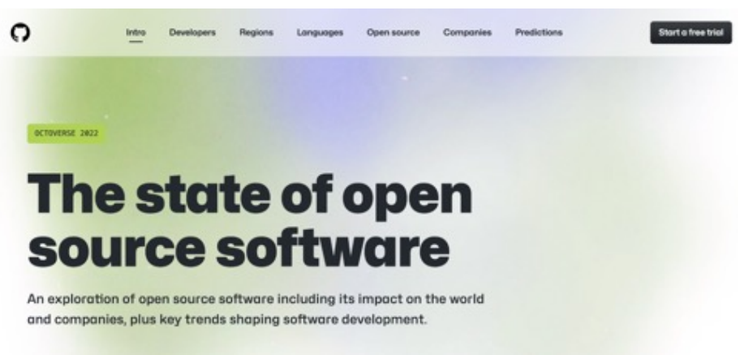
During the past five years, OSS development rose from approximately 35 percent to about 75 percent of organizations’ audited codebase.[1] As a case in point, by 2021, according to OpenUK, nine out of ten United Kingdom–based companies reported using OSS.[2] Organizations use OSS because it helps with cost savings, developer flexibility, and coding speed.

OSS usage creates opportunity for greater developer collaboration and allows coders to move quickly, because code libraries contain limitless amounts of prebuilt functionality and tooling resources. Having software components available on demand allows developers to leverage other developers’ work. Components are accessible and available from any location, and these elements are independent of individual manufacturers, making them attractive to start-ups and emerging technology players—or essentially to anyone who wants to build software quickly.

Although the benefit of incorporating third-party code is clear, OSS contributes to organizational risk. OSS is not regulated or overseen by a central authority and is publicly available. It can contain potential vulnerabilities, out-of-date code, and cyber exploits, which can expose organizations to cyberattacks. Most organizations seek to better understand and reduce their cyber and technology risks, but organizations recognize building and maintaining secure code is a vital cornerstone of any cybersecurity strategy.

6.6.2 GitHub - Octoverse 2022

https://octoverse.github.com/



6.7 OpenUK Survey 2023

You can find the survey at https://openuk.uk/stateofopen
It, like the content of this report is available for re-use with attribution.

6.8 Sponsors

We are grateful to our Report Sponsors GitHub and Civo and our in-kind sponsor FieldFisher and to OpenUK’s general Donors and Sponsors who can be found at https://openuk.uk/donors/ . Our Donors and Sponsors do not approve or disprove of our content.

6.9 Cover Photos

Cover photos are from our collaborator photographer Tiana Lea, who took 38 portraits for the OpenUK State of Open Exhibition sponsored by Arm[61], and has since taken a further almost 200 portraits at State of Open Con and our Honours list event. We are grateful to Arm for their continued support of this work and will continue to document the people forming the Open Technology community in this way.

Individuals on the front cover of Phase Two: Alexis Richardson, Founder & CEO, Weaveworks; Amanda Brock, CEO, OpenUK; Matt Barker, President & Co-Founder, Jetstack; Andrew Wafaa, Head of Open Source Program Office, Arm; Jennifer Barth, Chief Research Officer, OpenUK and Founder, Symmetry; Guy Podjarny, Founder, Snyk; James Governor, Analyst & Co-founder, RedMonk; Chris Eastham, Chief Legal Officer, Fieldfisher; Chris Howard, Lead Open Source Program Manager, EPAM; Amandine Le Pape, Co-founder and Guardian of the Matrix.org Foundation, COO and Co-founder, Element; Iain Mitchell KC, Honorary KC, OpenUK; Cheryl Hung, Director of Ecosystem, Cloud Native Foundation (Linux Foundation); Basil Cousins, Founder Open Forum Europe (retired); Paula Kennedy, Co-founder & COO, Syntasso; Dawn Foster, Director of Data Science, CHAOSS Project; Hiren Parekh, Director, WIT Consulting

6.10 In Memoriam; Basil Cousins, founder of Open Forum Europe

Basil Cousins was a stalwart of Open Technology in the UK and one of the first to recognise its value. Somewhat ironically, he was one of the founders of Open Forum Europe. He was recognised by OpenUK in its 2021 Honours List and participated in our State of Open Photo Exhibition, having a portrait taken and video made in January 2023[62]. Basil died in March 2023 and will be sadly missed.

61 https://openuk.uk/photoexhibition/
62 https://openuk.uk/soocon23/basil-cousins/

With thanks to our sponsors