DEEP DIVE: AI

FINAL REPORT

open source initiative®

EXECUTIVE SUMMARY

What does it mean for an AI system to be Open Source?

The Open Source community has shaped conversations during the emergence and rise of just about every technology we rely on: the internet, cryptography security, content rights management and patents on software. “The Open Source way” has also democratized technology, sped up collaboration, encouraged transparency, and generated community norms to encourage fairness. Now it’s time for the Open Source community to address one of the most powerful technologies of our age: artificial intelligence (AI), in a wide sense.

AI algorithms are growing more complex and pervasive every day. The traditional view of Open Source code implementing AI algorithms may not be sufficient to guarantee inspectability and replicability of the AI systems. It is time to address the question: What does it mean for an AI system to be Open Source? What policies are needed to both nurture innovation and protect individuals and society as a whole from harm?

After exploring these questions through our Deep Dive: AI discussions with experts in a variety of fields, we have identified three crucial areas where the OSI is compelled to give voice to the Open Source ethos:

1. **Open Datasets:** Vast amounts of data are required to train AI models; therefore, we must widen the availability of data (including images) lest we hamstring the development of AI technology that has inordinate potential to improve our lives. Society loses if we allow the extension of copyright laws to limit the use of data and images to the extent that only large organizations with deep pockets have resources to acquire access, as is already happening.

2. **Regulatory Guardrails, Not Roadblocks:** Modern-day AI systems, in general, are relatively immature, and we have much to learn about them. We are far too early on this learning curve to enact extreme regulations that will hamper innovation and advancement. Policy makers need to direct their efforts to creating essential guardrails that protect society without stifling innovation.

3. **Legal Frameworks for Ethical AI:** More than any other facet of society, AI practitioners themselves are most aware of and concerned about the ethical and moral ramifications of their work. We can help them by creating legal frameworks that allow them to collaborate and catalyze innovation but still help them sleep at night.

OSI will continue to seek input and opportunities to contribute the Open Source perspective on AI in a variety of forums in the coming year.

Stefano Maffulli
Executive director - Open Source Initiative

Table of Contents

Introduction

As the leading voice on the policies and principles of Open Source, the Open Source Initiative (OSI) helps build a world where the freedoms and opportunities of Open Source software can be enjoyed by all. A non-profit corporation with global scope, the OSI supports institutions and individuals working together to create communities of practice in which the healthy Open Source ecosystem thrives.

A goal of the OSI for 2022 was to explore the burgeoning field of artificial intelligence and to engage in discussions about what is acceptable for AI systems to be "Open Source." The topic is important because, for reasons explained later, AI ruptures the boundary between data and software and therefore raises the questions of if, when, and how the Open Source definition applies. Moreover, as AI knowledge and technology rapidly develop, the Open Source ecosystem has an important role to play in advocating for proactive AI policies and regulations that reflect and respect Open Source business, ethics and practice.

For these reasons, the OSI launched a three-part event called Deep Dive: AI to open a dialogue and explore the issues with some of the world's brightest minds in their respective fields.

1 FATHOM I : SIX PODCAST INTERVIEWS

Episode 1: Copyright, selfie monkeys, the hand of God
featuring Pamela Chestek
Episode 2: Solving for AI's black box problem
featuring Alek Tarkowski
Episode 3: When hackers take on AI: Sci-fi or the future?
featuring Connor Leahy
Episode 4: Building creative restrictions to curb AI abuse
featuring David Widder
Episode 5: Why Debian won't distribute AI models any time soon featuring Mo Zhou
Episode 6: How to secure AI systems
featuring Bruce Draper

2 FATHOM II FOUR PANEL DISCUSSIONS

Panel 1: Business perspective
featuring Stella Biederman, Astor Nummelin Carlberg, David Kanter, Sal Kimmich, and Alek Tarkowski
Panel 2: Society perspective
featuring Carlos Muñoz Ferrandis, Luis Villa, Kat Walsh, and Kit Walsh
Panel 3: Legal perspective
featuring Pamela Chestek, Danish Contractor, Adrin Jalal, and Jennifer Lee
Panel 4: Academia perspective
featuring Chris Albon, Ibrahim Haddad, Amy Heineike, and Mark Surman

3 FATHOM III IS THIS REPORT

which summarizes the discussions above and underscores what we've learned about the challenges and opportunities for the Open Source movement posed by AI.

We're using the term AI as a generic placeholder for different computer science disciplines, ranging from machine learning, computer vision and other. We're not entering into the debate of what constitutes "intelligence" and whether machines can ever be "intelligent."

A Brief Introduction to
ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

Before we recap what we've learned during Deep Dive: AI, let's do some level-setting with basic definitions and concepts.

EXECUTIVE SUMMARY

What is artificial intelligence?

For our purposes a simple explanation of artificial intelligence will do: Artificial intelligence is a subfield of computer science focused on developing computer systems that can perceive, synthesize and infer information. By learning from massive amounts of data, computers can make optimal decisions and decipher some kinds of problems radically faster and more accurately than humans can.

You may be old enough to remember when the "latest breaking news" in the field of artificial intelligence was when Deep Blue, an IBM chess program, beat world chess champion Garry Kasparov (1997). Twenty-five years later, artificial intelligence permeates many facets of our day-to-day lives. For example, you may have already used AI numerous times today through technologies such as smart voice assistants (Siri, Alexa, etc.), streaming apps like Netflix or Hulu, robotic vacuum cleaners, navigational tools such as Google Maps or Waze, facial recognition on your mobile phone, marketing and customer service chatbots, and all types of social media and retail applications that serve up ads and conversation streams for the products and topics in which you have previously expressed an interest.

One way to categorize AI is by the extent to which the computer systems are able to simulate human intelligence:

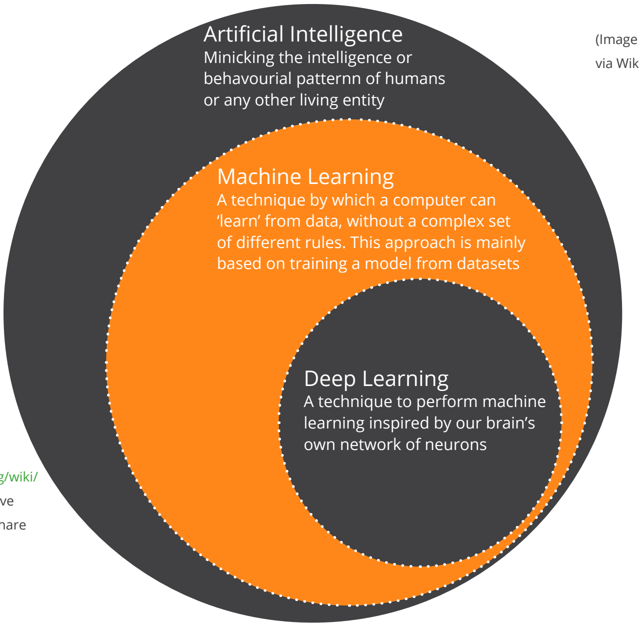
1. **Weak AI (also known as Artificial Narrow Intelligence, ANI):** In a weak AI system, the machine doesn't really think on its own; rather, machines are designed and trained to make necessary computations to accomplish a specific task. This is the category for all AI-based systems in use today.
2. **Strong AI:** Theoretically, in a Strong AI system, computers think and behave exactly the way that people do (Artificial General Intelligence, AGI), or perhaps even smarter and faster than humans (Artificial Super Intelligence, ASI).

What's the difference between AI and Machine Learning (ML)?

The research fields within AI are numerous and include computer vision, natural language processing, machine learning and deep learning. Machine learning is a subset of AI that gives computers "the ability to learn, adapt and improve from experience without being explicitly programmed." ML systems use algorithms and statistical models to analyze vast amounts of data; by interpreting and drawing inferences from patterns in the training data, the systems are able to recognize patterns and make predictions about outcomes when pieces of new input data arrive.

A very simple example of machine learning would be training a computer to predict a person's income based on their years of education. In this scenario, humans provide the computer with a set of training data (a table of both inputs and associated outputs–in this case, people's years of education and their associated income) and a linear regression model. The machine computes a regression line through the training data and then uses the regression line to predict income (X) when given a person's years of education (X).

There are three categories of machine learning (supervised, unsupervised and reinforcement learning) and many types of machine learning models that enable more advanced (and now common) applications, such as image recognition, automated language translation and traffic alerts.
The key aspect to classical machine learning is that humans tell the computer what features to look for to best predict the outcomes.



Artificial Intelligence
Minicking the intelligence or behavourial patternn of humans or any other living entity

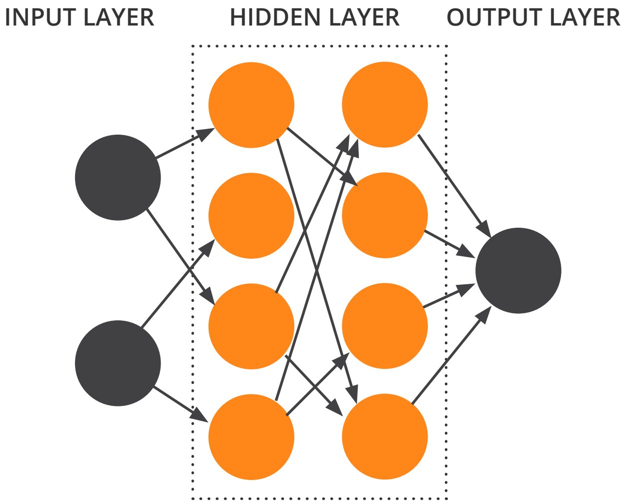
Machine Learning
A technique by which a computer can 'learn' from data, without a complex set of different rules. This approach is mainly based on training a model from datasets

Deep Learning
A technique to perform machine learning inspired by our brain's own network of neurons

(Image Source: https://commons.wikimedia.org/wiki/File:AI-ML-DL.svg, Creative Commons Attribution-Share Alike 4.0 International)

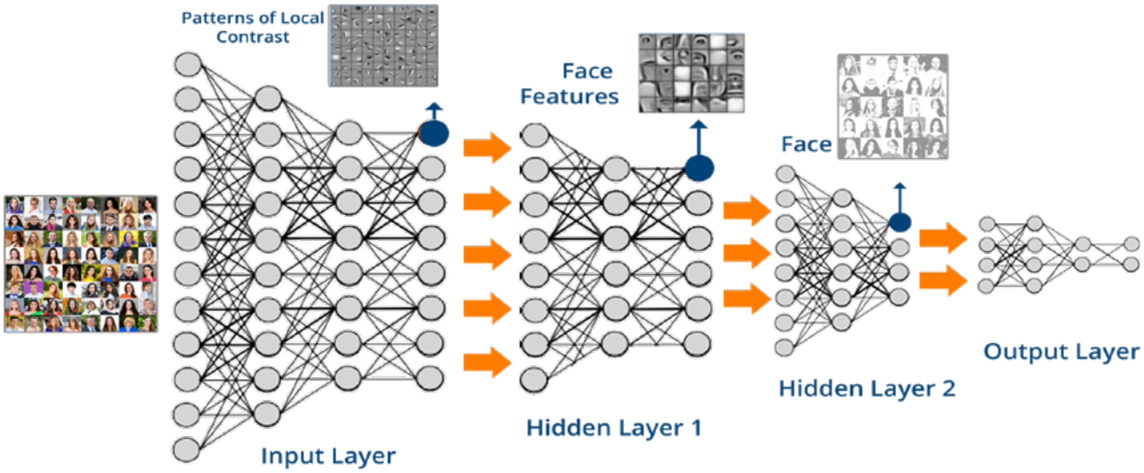
What is deep learning?

Deep learning is an advanced subset of machine learning. In contrast to classical machine learning, deep learning systems are capable of figuring out by themselves the best features to focus on to most accurately predict outcomes. Deep learning systems can ingest unstructured data in its raw form (e.g. text, images) and automatically determine the hierarchy of features which distinguish different categories of data from one another.
Deep learning comprises a brain-like layering of algorithms (referred to as artificial neural networks, or ANN). The middle layers of these networks are called hidden layers because their values aren't observable in the training set; the hidden layers are calculated values used by the network to do its "magic." If a system has more than two layers between the input layer and the output layer, it is considered to be a deep learning system, and the more hidden layers a network has, the deeper it is.



(Image Source: https://commons.wikimedia.org/wiki/File:Neural_network_explain.png, TseKiChun, CC BY-SA 4.0, via Wikimedia Commons)

In the case of facial recognition, for instance, the first layers of the neural network take a set of photographs and determine patterns of contrast. The next layers may identify facial features. And the final layers may apply facial features to face templates, ultimately being able to identify full, unique faces.



(Image source: https://www.edureka.co/blog/what-is-deep-learning)

Deep learning requires much more data than a traditional machine learning algorithm to function properly. Whereas machine learning may work with merely a thousand data points, deep learning oftentimes requires millions. Due to its complex multi-layer structure, a deep learning system needs a large dataset to eliminate fluctuations and make high-quality interpretations.

For more information, see **The Deep Learning Book.**

Sources of the information and illustrations in this section:

https://www.computerworld.com/article/2906336/what-is-artificial-intelligence.html

https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/?sh=6b82b75b4f5d

https://www.ibm.com/cloud/learn/what-is-artificial-intelligence

https://levity.ai/blog/difference-machine-learning-deep-learning#:~:text=Machine%20Learning%20means%20computers%20learning,documents%2C%20images%2C%20and%20text.

https://www.softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/

https://builtin.com/artificial-intelligence/artificial-intelligence-future

https://builtin.com/artificial-intelligence/examples-ai-in-industry

https://insights.daffodilsw.com/blog/10-uses-of-artificial-intelligence-in-day-to-day-life

https://www.edureka.co/blog/types-of-artificial-intelligence/

https://www.edureka.co/blog/what-is-machine-learning/

https://www.edureka.co/blog/what-is-deep-learning

The Power of AI

WHAT IS THERE TO FEAR?

One clear sentiment resonated throughout our discussions in Fathom I and Fathom II of Deep Dive AI—AI is forever changing our world. AI has already impacted virtually every major industry. The pace of research and innovation in the past few years is staggering; in fact, what was considered cutting-edge three years ago is not even being used today because we've progressed so much. But, we've only scratched the surface. We're still in the very early days of figuring out useful applications.

"AI is also the most powerful technology of our time to improve our lives, to allow us to address tons of problems that we are currently facing....If we could resolve scientific bottlenecks faster, if we could just develop cures faster, if we could just do all these things faster and more efficiently, we could improve society immeasurably.... That would make the world so much better, more than almost anything else in the history of mankind."

Connor Leahy (CL 31:52)

"I received...a legal document that was written in French and had simply been put through a machine translator for the English translation....We're to a point now where we rely on the machine learning; I think it is probably always the first step in any translation at this point....We're still at that very early stage right now of, 'Yeah, I can get the gist of it. It's not great,' but we will be getting to a point where it's just going to be part of the ordinary fabric of our lives to rely on all of this machine-generated content, or decision-making by machines."

Pamela Chestek (PC 28:19)

"It's staggering the pace and diversity of the research and innovation that's going on in AI right now....It's also very, very early days for us figuring out how you make this stuff useful in people's day-to-day life."

Amy Heineike (Academia 07:17)

"I could name a bunch of stuff, like recurrent neural networks or convolution neural networks, which were cutting edge five years ago and now totally not cutting edge at all. And no one uses it in production....[At that pace it] is a really hard thing for societies and particularly governments to catch up on."

Chris Albon (Academia 26:35)

And what of the risks?

Another shared sentiment among our contributors is that AI is a tool that can be used for both good and bad. We've been preconditioned by many decades of "mythologizing" in science fiction books and movies to fear AI, just as other novel technologies like fire, electricity, steam engines and airplanes have been feared in the past. We certainly don't want to stymie innovation based solely on our fear of the unknown, but we also need to recognize legitimate risks and guard against societal harm. Today's practical applications of AI have already raised serious concerns about the need to protect society from abusive or unethical AI practices, including biased decisions, surveillance of citizens, and invasions of privacy, to name a few.

Deep Fake AI technologies, for example, have been used not only to create comical parodies but also to falsify news, to influence elections, and to create fake pornography to damage the reputations of innocent women. AI-powered drones can deliver life-saving organs to transplant recipients in rural areas or to commit acts of mass violence against crowds of people. Surveillance systems can be used to apprehend terrorists or to control dissent against abusive governments. Organizations have already been successfully sued for using AI systems that introduced racial, ethnic, gender and other types of bias in decisions such as who receives mortgage loans, job offers, or parole.

"There are a lot of people concerned about the advancement that is happening and what kind of potential danger that poses to society in general."

Ibrahim Haddad (Academia 46:26)

"Of course, you should be concerned when you build a new technology that has unprecedented capabilities."

(CL 14:27)

"Today's AI systems have incredible sophistication that did not exist two to three years ago. Extrapolate that progress...what could be possible? Could a system manipulate other systems? Suppose a wealthy corporation has access to a system that's 1,000 times smarter than any other human, better at hacking, better at planning, better at propaganda. It can generate images, videos, voices and impersonate anyone."

(CL 21:2, 24:09)

"Say we have some really big corporation...that builds the biggest system of this kind ever—something so powerful that it's smarter than humans, it runs a million times faster, it's read all the books in history, it can do perfect IMO gold medal of mathematics—and then you give it some goal like maximum profit. What will such a system do? I think if you meditate on that question a bit, the obvious things are not always good, or even mostly not good."

Connor Leahy (CL 21:22)

"I like a term that sometimes is used alternatively to 'artificial intelligence,' which is 'automated decision making.' Basically, automated decision making says there are situations where humans no longer decide about you."

Alek Tarkowski (AT 02:26)

"I remember there was one Twitter bot that turned racist in about eight hours. In almost no time it was creating racist slurs and they had to take it down...it was just a disaster. That gives me great pause when I see these tools being used prematurely, being relied on in ways that are harmful to us."

Pamela Chestek (PC 32:56)

"One of the myths that I encounter, not just for AI but any algorithmic decision-making tool, is the idea that the machine is neutral and wise....We all cringe when we hear that, but it does require a lot of educating of lawmakers about how [algorithmic decision-making tools] embody biases, not just of the programmers but also of the data and assumptions that go into them. Which leads me to something else that's a little bit different about AI, which is the way that its development is so dependent on massive data collection. It raises a lot of new privacy related concerns, especially in terms of data that reflects private information about individuals, information like pictures of my face that I put on Flickr and never thought once that it would be used to train a facial recognition system that police would use to imprison people. And another element that I think is pretty different is the explainability piece—how many different artifacts of machine learning development you need in order to have a shot at being able to explain how it is arriving at those conclusions."

Kit Walsh (Society 07:31)

"Financial and reputational harms are definitely just a few of so many harms that arise from non-consensual data collection. Whether or not that's used by companies or individuals...those harms can lead to stalking, domestic violence, police violence.... Data can be used in so many harmful ways that can seriously lead to life or death consequences. Data could be used to track people who seek abortions, for example...And I think that's something that companies, governments and individuals all really need to be thinking very carefully about."

Jennifer Lee (Legal 24:29)

The Incredible Importance
OF DATA SETS

As Connor Leahy described in his podcast session, three primary components are required to build an AI system: data, engineering, and compute (hardware) resources. Our discussions in Deep Dive AI revealed how the resources required in each of these areas can severely limit the entities that are able to harness the power of AI. Take data, for example—massive amounts of data are required to train AI models. For text training, explained Connor, "you need truly stupendous amounts of text. A rule of thumb is you want a terabyte of raw text." Even more data is required for an image-based AI application like facial recognition. One popular dataset called ImageNet contains more than 1 million images. The quantity of data required to train AI systems presents a problem. As Mo Zhou asked in his podcast, "Who can collect such a large-scale data set? Only large corporations have enough funds to do this. It is very, very difficult for any person to do this." (MZ 15:27)

AI also requires specialized hardware that is extremely expensive, difficult to set up, and often requires proprietary drivers. "Compute is actually the biggest bottleneck," said Leahy. "The amount of computation that goes into building something like GPT-3 is massive. And it's not like you can run it on your CPU. You need massive clusters of GPUs, all interconnected. This is high-end, supercomputing-grade hardware, which is very expensive and quite tricky to use." (CL 06:49) Mo adds that these high-end systems are the only way to run large scale AI models in a timely way. "The speed issue is critical, because if you train a neural network on a CPU, it may require several years, but if you've got GPUs, only several hours." (MZ 27:37)

Because of the vast data sets and expensive (and often proprietary) hardware required to develop AI models, large corporations are the primary players in the world of AI today. They can afford to conduct their own proprietary research and development; they can afford to enter into commercial agreements for licensed access to data sources; and they can afford the expensive computational hardware that is far out of reach of others. Some government, academic and research institutions are capable of participating to a degree, but smaller-scale organizations and individuals are largely excluded from AI development. Fortunately, we're beginning to see changes with respect to availability of datasets and, more recently, models are increasingly shared with fairly permissive terms and conditions, thanks to the contributions of organizations such as EleutherAI, Hugging Face, Hippo AI Foundation, ML Commons and more.

> *"Whether you're interested in text generation or text image modeling or reinforcement learning for playing games, the overwhelming majority of the research in this field is controlled by very large tech companies and a very small number of them globally. They have a lot of money and resources and influence to be able to … pump out this research very quickly….Roughly speaking, if you have twice as much money, you can finish the model twice as quickly… I've trained a 20 billion parameter language model called GPT-NeoX, and that took me about three months. If I had enough money, I could have done that in three weeks instead of three months. The difference there is solely about the number of GPUs I can afford to pay for. And we know that DALL-E, for example, that OpenAI developed was trained in less than a month, and that's really a statement about resources, not about the AI itself. It's a statement about the fact that OpenAI actually had the resources to go do that. Someone else could have trained the same model in a year."*

Stella Biderman (Business 27:37)

Provenance of Datasets:

Do copyrights apply?

As important as datasets are to AI advancement, the topic of datasets can be quite controversial. One of the issues at play is provenance. Where does all the data come from? Building a dataset of sufficient size to train an AI model typically requires scraping data or images from the internet, publicly available databases, or proprietary sources. As we've already mentioned, careful consideration must be given to the data sources to avoid inadvertent statistical bias in the data used to train AI models. Other provenance issues arise from legal concerns, especially copyright, licensing and rights to privacy. Can copyrighted or licensed text, images or code be freely used to develop training datasets? Are datasets themselves copyrightable or licensable? Does use of my personal data and images without my express consent for the purposes of building a data set violate my personal right to privacy?

We explored these questions at length in many of the segments of Deep Dive AI. Here are a few excerpts to illustrate the challenges:

"I don't know where all of the data, where all of these photos are coming from that are used for training, but the subject matter that is being used for training is subject to copyright....Is all of that data allowed to be used? Was it used with permission? If not, what does that mean about the model? So if I've used data that I shouldn't have used, I didn't have permission to use, does that taint my model? (PC 22:25) *"I think everybody's just kind of speculating and assuming, but when the courts get to this, I think there will be a lot of unpacking ... we just don't have a clue where and what pieces of this copyright will apply to."*

Pamela Chesteck (Legal 06:59)

"...we're also seeing lawsuits around GitHub Copilot *[software code generator], saying 'Hey, you're taking all my code and then serving it to people basically as a product'...We're seeing* DALL-E *and other generative art things literally sometimes displaying the watermark because they had scraped all these stock photography sites that have the watermark....I definitely think one of the biggest things that would be interesting to see is more clarity from the legal and licensing perspective around the data that's being used.... So maybe they take Wikipedia and have one license for the content, and then they grab everything from GitHub, which has a different license, and then they go to DeviantArt...which has a different license for every single artist...And the end database is actually this hodgepodge of a bunch of different licenses that have been thrown together....I would love to see a really permissive data license specifically around ML/AI."*

Chris Albon (Academia 54:53)

"Historically speaking, machine learning researchers tend to think that if they go out and collect or reprocess or repackage data, it's now its own thing and they can license it however they want and there is no provenance before that....That is just simply not true.... And now we're in a really awkward spot where there are a lot of very widely used datasets that explicitly have falsified provenance that are used by thousands or even more researchers all the time ... and [there is] no real ability to either prevent that from happening in the future or undo that. Building the correct documentation of even a relatively modest ...dataset is an exceptionally large amount of work and it's not something that organizations, the companies, have the resources to do or really care about doing."

Stella Biderman (Business 31:20)

"At Open Future, we've been looking at a very specific case.... It's a story that's by now almost 10 years old. It's a story of huge numbers of photographs of people being used to build datasets with which facial recognition models and technologies are built. There's a famous or infamous data set called MegaFace*, which has 3 million photographs packaged into a tool that's an industry standard for benchmarking for deploying new solutions in face recognition training. It's also a system that's quite controversial. When you dig deep inside, it's not entirely clear consent was given. A lot of people say that even though the formal rules of licensing of Creative Commons licensing were met, they see some problems. The uses are unexpected, risky. When you look at the list of users of this data set, you suddenly see the military industry, the surveillance industry, and people really have some disconnect between the ideas they had in mind when sharing their photos. Yes, they agreed that this will be publicly shared, but they usually imagined this as a form of participation in a positive internet culture. Then they find out about these unexpected, even scary use cases..."*

Alex Tarkowski (AT 41:47)

"When we think about developing new tools and collecting information, we really need to be thinking about the end impacts on who's going to be the most harmed. Although technology has advanced tremendously over time, surveillance isn't a new concept. People have always been surveilled over history and new ways of gathering enormous amounts of data just make it really easier to target historically marginalized communities.... The surveillance laws that we're working on, the privacy laws that we're working on, and the AI regulation laws that we're working on–they impact each other."

Jennifer Lee (Legal 10:29)

“One of my concerns going into the future is that not all of the databases that you’ll be pulling from are necessarily static databases. The data itself may change and drift over time, and you need to know whether or not that’s still trustable as a signal that you need for your algorithm.”

Sal Kimmich (Business 01:03:45)

“There’s a great point...datasets do mutate over time...There’s a cost to that, which is it makes research and a lot of other things very difficult. One of the things that I’ve looked at with some trepidation is that GDPR essentially makes almost every dataset that contains personal data conditional. So every single dataset that contains anything that might conceivably be personal information or can be used to derive personal information is not a static dataset.”

David Kanter (Business 01:05:08)

Copyright protections for data, images and databases vary from country to country. In the US, numerous battles are being waged in the court system, and “the jury is still out” on how copyright laws will impact the collection of AI datasets or if collection of data is considered “fair use,” even if it’s done for profit. A couple of years ago, the European Union enacted a Directive on Copyright in the Digital Single Market (CDSM) that provided an exception for text and data mining (TDM) for the benefit of non-commercial research organizations and cultural heritage institutions. The UK is expected to enact legislation soon that would extend TDM permissions for all uses, in order to encourage and support AI development.

We’ll further explore regulatory and licensing issues later in the section titled “How Do We Manage the Risk?”

Data is crucial to train AI models, and if we want to enable innovation and improvement of AI models, then we need to give AI practitioners more freedom to access this data.

If AI practitioners need to ask for permissions to access data to build their training datasets, datasets will be smaller and the model will be less accurate; conversely, if data is considered “fair use” and “open” for the purposes of building AI training datasets, AI practitioners will have more data at their disposal, and training of models can be more effective. Although software developers, writers, artists, and private citizens raise valid concerns about the ownership of their work and private information, those concerns need to be balanced with the needs of society as a whole. As Pamela Chestek dared to ask, “Is it possible that not having copyright protection for these works actually is the best solution?” (PC 18:29) Society as a whole will lose if we ‘fence in’ all this data and make it so that only the large organizations have access. The same is true for the hardware required to run the models. Going forward, OSI will be looking into what can be done to widen the availability of data in order to get better AI systems while mitigating new types of risks.

The Mysterious “Black Boxes”

OF AI/ML MODELS

As depicted earlier in our brief description of deep learning, many AI models consist of “hidden layers” of algorithms that identify features and patterns in data and images without humans having to provide specific directions to the computer. Because many AI computations occur within the computer and essentially without human control and observability, AI models are often described as mysterious “black boxes.” We can’t always explain why an AI model decides what it decides. Therefore, we can’t prove that the machine is neutral and wise. This is often referred to as the explainability or interpretability problem.

> *“There are some important ways in which [AI] is different [from other technologies]...One, in particular, is AI explainability. More complex ML algorithms and many neural networks...are currently somewhat inscrutable.”*
>
> David Kanter (Business 12:28)

> *“We’re at the point where we have these super complicated systems...and we just have no idea what’s going on inside of them....It’s not possible currently that we say ‘Oh, we see a failure case in our model. That’s not good’ and then we go into the model and fix it. We can’t currently do this.”*
>
> Connor Leahy (CL 42:01)

> *“The issues [with AI] are the same as with all the automation, but the ways of addressing them are a lot harder, because there’s this beautiful symbol of the black box that hides things inside.... complexity [of AI systems] basically makes them much harder to analyze, assess their impact, and so on.”*
>
> Alek Tarkowski (AT 06:58)

> *“I think it’s very scary to think of models ...suddenly being given the power to decide who gets hired or doesn’t, and that could be biased in some deep way. How would we even know that, if we don’t have access? ... If we build applications on top of this when we don’t know what it’s doing, how do we check that it’s doing what we even want it to do? We might have good intentions and just not have realized that it was deeply flawed, or we might have bad intentions...”*
>
> Amy Heineike (Academia 36:08)

> *“I think one fundamental difference between AI and software (or AI and humans) is that we, humans, are very causal creatures. We understand those causal relationships. If you tell me why you make a decision, then it’s much, much easier and much more intuitive for me and a society to decide if it was fair or not.... Whereas when we talk about AI systems, in most cases we can’t necessarily interpret them. We don’t necessarily have tools...that would give us the explanations on why a system made a decision the way it did.”*
>
> Adrin Jalali (Legal 37:17)

Explainability is just one of many AI challenges yet to be solved. Connor Leahy described several more (CL 32:50+) :

- **The Alignment Problem:** How do we get an AI system to do what we want and align with countless unspoken values and innumerable hypothetical scenarios that may occur? For example, if I train a robot to bring me a cup of coffee, how do I ensure it won’t run over grandma and my cat in the process?

- **The Stop Button Problem:** How do we make robots that are entirely indifferent to being shut off?

- **The AI Security Problem:** How do we keep AI systems safe from hackers?

AI technology is relatively young, and so it’s not surprising that the list of AI problems to be solved is substantial; nevertheless, some experts are optimistic, believing that we’ll figure out the solutions to these problems eventually.

How Do We Manage

THE RISKS OF AI

Some would argue that AI is different from other technologies and poses unique quandaries regarding ethics and responsible use. Others would say that AI isn't all that different from radical technologies that have come before–it's just that society as a whole and policy makers don't understand it well enough yet to regulate it.

As a group, our contributors did agree on two things: (1) some combination of regulation, licensing, and community norms is needed, and (2) we have huge hurdles to overcome. One of the biggest hurdles is the complexity of the technology; currently, there are relatively few persons who are willing and able to educate policy makers and other key stakeholders on the technology and its legal, business and societal implications. In addition, societal and government values vary widely throughout the world; what one government deems an infringement of human rights, another may consider fair use.

To date, the European Union is clearly leading the way in proactively addressing the fair and ethical use of AI. Alek Tarkowski briefly summarized recent EU legislative actions for us during his podcast:

> *[Europe] has the recently adopted Digital Markets Act and Digital Services Act, which regulate platforms.... And then we have the AI Act....This is exactly the right moment to have a conversation about AI regulation, and I'm happy it's happening right now.... Yet [the AI Act] doesn't really cover all regulatory issues related to AI; it really focuses on two issues: what kind of AI uses or technologies are so dangerous that they should be outright banned, and what kinds of technologies, or contexts in which they are used, are high risk–risky to the extent that you need to regulate their use."*
>
> (AT 09:39)

Alek explained that banned uses include subliminal distortion, social scoring, biometric identification, and vulnerability exploitation. High-risk uses may include applications in such arenas as employment/HR, predictive policing, migration, justice and democracy.

In the US, AI policy formation is still relatively formative. As mentioned above, stakeholders are still waiting on the courts to definitively weigh in on the applicability of copyright protection to data and images, datasets, models, and outputs. In December 2020, President Donald J. Trump issued Executive Order 13960 addressing the proper use of AI by the federal government. In early October of this year, the US White House Office of Science and Technology Policy (OSTP) released a "Blueprint for an AI Bill of Rights," a non-binding set of recommendations for using AI in ways that won't compromise the safety of Americans. To date, federal AI legislation has not been passed in the US, but numerous states have issued legislation of various forms.

> *"And [AI] is sort of like any technology: it's going to start out as arcane black magic until at one point it's prosaic. Flight is a great example of that. I don't know anyone who takes a flight on a commercial airliner and says 'I need to study what plane I'm on to figure out if I'm gonna die because it's going to crash.' Now that was a reasonable thing to do in the 1910s. But today, flight is just this amazingly magical and beautiful thing that's in the background for almost everyone, and one day I hope AI will get there."*
>
> David Kanter (Business 16:06)

> *"...[I] think that artificial intelligence is the most important technology of our time. And as it becomes more and more powerful, and it can do more and more tasks, it will become a more and more powerful and dominant force in our society. It's very important to understand this technology. That's one of the reasons I now have a startup where we work on researching safety of AI systems, how to make them more reliable, how to make them safer, how to make them not do things we don't want them to do, which is a big problem with AI and only will become more of a big problem.... We work on interpretability research. We try to take the inner parts of these networks, decompose them into understandable bits, and then see where failure modes come from. How can we edit these things? How can we manipulate them? How can we test them for safety features, and so on....There's a massive amount of promise here that we're just starting to unearth, and I think there's a real reason for optimism."*
>
> Connor Leahy (45:03)

An important revelation from Deep Dive:

AI is that modern-day AI systems, in general, are relatively immature. If "AI Enlightenment" can be described as an era, we're clearly still in the very early years. We're just now trying to fix bias in datasets, improve the output accuracy of algorithms, and understand how the decisions actually get made by neural networks. Can we look inside a model and fix problems? Connor Leahy described a way to look inside a model and make it believe that the Eiffel Tower is in Rome instead of Paris. It can be done in theory, but we don't yet have the tools to do it at scale. There are very few non-AI practitioners who understand exactly what's going on inside AI systems and how different the functioning and impact of AI is from that of traditional software. This is something that regulators, in particular, need to understand.

"The conversation about AI regulation is just starting. People have been talking about it for years, but in terms of policy—actual laws around regulation—at least in the United States, that's something we're getting to just broadly."

Jennifer Lee (Legal 24:29)

"[W]e also have to pay attention to how these systems are ultimately used, rather than just how they are built...."

David Widder (DW 06:57)

"Who actually gets to determine what's harmful? Is it regulators? Is it the developers? Is it people who are using these technologies? It's typically not the people who are actually experiencing the harm (Legal 39:45)*....I think regulation can go a long way in mitigating some of those harms. Requiring transparency and accountability for these sorts of technologies is really important. But ultimately, the question I ask when I look at proposals is 'how is power distributed?'...It's often not going to be completely solved by regulation, but it's a step in the right direction* (Legal 41:46) *....When we're looking to figure out solutions on how to best regulate AI, I think it needs to be a more collaborative process, really bringing in the communities that are impacted.* (Legal 01:01:10) *"*

Jennifer Lee

"Our rules are not just meant to protect citizens. I think there is, rightly so, a lot of debate that basically asks the questions, 'How are we going to be safe?' 'How will our rights be protected?' Yet there are more questions that need to be asked... which are, 'How will we make this technology productive in a way that is at the same time sensible, reasonable, and sustainable?' ... 'How can this technology be used well?' Because I think this is something that can easily become forgotten when we only talk about risks—that there are positive uses of these technologies. There's a huge promise that you can find ways of using data to the public benefit, but it requires smart regulation, and supportive policies."

Alex Tarkowski (AT 18:56)

"Regulation has an important role to play in AI technology, especially where high-risk applications can be better defined. (Legal 01:04:36) *...Now, if you were to say, regulation can solve all of this, I would be suspicious of that. I argue that its 'terms of use' (depending on the capabilities of a technology) that one should really be thinking about.* (Legal 01:06:56) *And I think norms around releasing AI and software need to adapt to what we are seeing happen in the real world. In my view, it's unsustainable to release AI systems without restrictions.* (Legal 01:13:43) *But I don't think there are any easy answers. I don't think regulation can solve everything. I don't think licenses can solve everything. I don't think transparency can solve everything. It really has to be a considered effort addressing all of these different touch points."*

Danish Contractor (Legal 01:29:18)

"The number of times I've had to explain to people that license terms are most effective against large entities with legal departments and not against individual bad actors! That is a very long-time discussion and relevant here, because some of the kinds of things that we're trying to regulate in the AI space are very much the things that governments do, that large commercial entities do, and for them a license may be a very effective regulatory regime because their lawyers actually do read the things. But if you're trying to say, 'Don't use this image generator for porn or you violate the license, [licenses aren't very effective.]"

Luis Villa (Society 45:13)

"If there is going to be a policy enactment, it has to be categorical. Categorically, right now I do not think we have anything in place in order to be able to identify for whom the intent belongs to for negative secondary effects of a dataset produced by, or an algorithm produced by, an artificial agent.

Sal Kimmich (Business 22:51)

"Anticipating harms and limitations of technology is an important aspect that I think all developers should consider, especially in AI, just because of how the [models] can be repurposed into larger software systems that they were not originally designed for. ... I think even at the point of release, if developers are aware of certain limitations and restrictions, I think those should be made part of terms of use, because that only just gives enforceable mechanisms for preventing harm. Otherwise, if you don't put that in your terms of use as a model creator, as a developer, you don't even have rights to enforce anything."

Danish Contractor (Legal 43:34)

Deep Dive: AI has delivered some good news with respect to the issue of ethical use of AI:

Many practitioners of AI, and especially the most visible ones that are working with large models, are fully aware of the danger of misuse of their models, and they are being extra careful, thinking about how to liberate their thoughts and their code, what is safe to share, and whether there is a way to limit the usage. That sounds like a reasonable and safe approach, right? Here's the quandary: Ultimately, is limiting the use of AI models the best approach for society? Will "AI in the hands of a few" stifle scientific advancement and grassroots innovation?

The Role of Open Source
AND THE VALUE OF DEMOCRATIZING AI

What does Open Source have to do with AI? For starters, AI researchers, individual developers, Open Source communities, and even private corporations are already "open sourcing AI," making available datasets, models and tools, free for others to use. In fact, you can peruse the websites of many of our contributors' organizations and find a wide range of freely available AI resources at your disposal.

But as we have summarized above, concerns about ethical and fair use of AI abound. The OSI is trying to understand how we can help. In the realm of software development, OSI has provided a way to identify the basic needs for developers and citizens to enjoy life in the digital space. We'd like to contribute to the achievement of the same balance with regard to AI. Over the course of our Deep Dive: AI program, we discussed at length the role that Open Source plays in supporting AI innovation while protecting society from harm. We asked our contributors what value OSI can bring to the table in that regard, and they shared many valuable insights. Here's a selection of their thoughts:

"As we begin to talk about ethical AI, if we let companies only drive this conversation, and we don't look to Open Source, and we don't look to public sector organizations, then I think we're going to get a very particular idea of what ethical AI is and what kind of problems there are, that is going to be driven by the interests of big companies....I think the Open Source strength in this area is what it's always been, which is the broad diversity of communities that can set their own norms, that can refashion these norms as a way to experiment on what ethical AI might mean, in a way that is not dependent on the for-profit context..."

David Widder (DW 43:38)

"[I'm concerned about] regulations hampering open development by creating standards that are only able to be met by some of the largest companies and stopping other development from happening. I can see those processes all the time...creating standards that only the largest corporations can meet. That's the thing I don't want to happen." (Society 29:51) *"I would love to see us basically create rules that would keep that openness, keep that generative ability which we have seen with other technologies like the printing press and software development in general. If we can find a way to balance regulation of the harms with keeping that system open, AI in general has a great possibility to enable more people to create and participate in creation."*

Kat Walsh (Society 01:19:48)

"I come from a tradition of open content activism of Creative Commons, which borrows heavily on the Open Source philosophy and methods and basically deploys this most basic but also extremely functional tool, which is an open license. And the licensing mechanism has really been solving a lot of issues around access, around sharing of content, around intellectual property of code. But I think this is why the debate about AI is interesting; at some point, it shows the limits of just saying 'Open it up.' We know that a lot of the technological stack of AI systems is open—it's based on open code—and that doesn't solve any of the problems of the black boxes of AI, and uses that are unexpected, and possibly lead to societal harms. I think we need to take the spirit of Open Source, of openness, but really look for some new solutions."

Alek Tarkowski (AT 27:23)

"The OSI is very clear that we do not discriminate against commercial or non-commercial uses. There's a reason for that: the line-drawing gets very difficult. And the good and evil question is just insolvable, as far as I'm concerned. And so we have to take this position of 'We're not making value judgments on how this stuff is used...Knowing in particular that there are problems with models...and how harmful that's going to be, there is a big part of me that wishes that we could say, 'No, it's not consistent with our belief system that these should be used,' but...we have always been very clear that OSI software can be used for evil....I don't see a way to draw a different line for models."

Pamela Chestek (PC 27:29)

“[Open source] brings not only a lot of access but also a lot of speed to the development itself.”

Astor Nummelin Carlberg (Business 09:19)

“One of the things that I think about and very strongly believe in at MLCommons is the ability to create open data, open metrics and all of these things to help democratize AI. That’s my mission.... How do we take something that people today see as magic and make it ordinary magic that suffuses our daily lives and in a way that doesn’t violate the expectations of individuals (Business 01:27:07)*....How can we bring some of [Amazon’s internet AI] magic into the hands of a mom-and-pop shop and what does it take to get there? There’s so much friction we need to eliminate in terms of making things easier to train, easier to deploy, et cetera.”*

David Kanter (Business 01:28:08)

“I think there’s massive value from Open Source that lends itself to the domain of AI in general and data as well, and that can be summarized in one word, as everybody mentioned: transparency.... [We] try to look at four different challenges with that spec. The first one is ensuring fairness. So we need to be able to have Open Source tools, methods, libraries, whatever the case is, that would allow us to detect any kind of bias, whether it’s in the data sets or the models....The second aspect is what’s referred to as robustness, which are basically methods, libraries and tools that allow us to detect if there has been any tampering with the data sets and the models. The third aspect is explainability...And the fourth aspect is lineage, and that applies to both data and models as well. Basically it’s understanding the origin of the methods, models and the data sets, any changes that were done to them by whom....I think this would be a great [way] to address the general issue of ethical AI by addressing the smaller subsets of challenges in relation to fairness, robustness, explainability, and through the Open Source methodology of collaborative work, openness and transparency.”

Ibrahim Haddad (Academia 38:11)

“When it comes to harm, to me it’s a lot more a matter of regulation than it is of licenses. We shouldn’t be doing certain things no matter whether the person releasing that software or that model was okay with us doing that or not.”

Adrin Jalali (Legal 38:23)

“So we are now in this very moment fighting this policy battle for openness versus closed, or ‘openness equals democratizing access to machine learning’ versus ‘closed machine learning is safe machine learning,’ right? This is the main policy debate we are having right now....What do we do right now if we don’t have regulation but at the same time we want to keep promoting open access and responsible review? (Society 35:16)*...Open RAIL was basically the response to this issue. Are RAIL’s the silver bullet? No, of course not. They are just another proposal by the AI community which should be improved by taking a collaborative approach.”*

Carlos Muñoz Ferrandis (Society 39:08)

“So even within the dictates of Open Source and free software, I think that we can think a little bit more creatively about how we can build restrictions about what uses we think are immoral or unethical into our software, and not see it as black and white. It’s not always going to be we stop all misuse, or we just don’t stop any, but there are shades of gray in the middle.”

David Widder (DW 16:05)

“Who knows how we get to ’trustworthy’? But one piece is transparency and building incentives for it at this point....Let’s start to figure out what that can look like in practice...What does fairness or at least fairness in design look like in practice? And we have decades to go in trying to figure that out....Look at what kind of licenses or other tools might let us explore those topics faster than law...I mean, it’s very vague at this point, but I do think we should take up the spirit that Open Source had 25 years ago and look at how it would go faster than the law in prototyping some of these questions.”

Mark Surman (Academia 32:32)

Contributors were “preaching to the choir” when they advocated all the ways that an Open Source perspective could benefit the field of AI.

Music to our ears were statements about how Open Source speeds up collaboration, democratizes technology, encourages transparency and collaboration, and generates community norms to encourage fairness. All true. But also true is how murky the waters are when we try to define “Open Source AI.” Does the Open Source definition apply to AI and ML as it is currently written?

What about Open Source licenses: are they sufficient or insufficient to apply to AI? Can Open Source principles make AI safe enough to “run in the wild”? Does Open Source AI warrant its own unique regulatory approach? What role, if any, does Open Source play in defining and enforcing AI ethics? The questions are many; the answers are few. But we’ll be working on that!

What’s Next?

ADDRESSING OUR MOST PRESSING ISSUES

We are indebted to our esteemed group of contributors who provided their time, professional expertise and wise insights to Deep Dive: AI. Our sessions were invaluable in providing OSI and our constituents with “a bird’s eye view” of the issues surrounding one of the most powerful technologies at our disposal today.

“We’re in this incredible time of innovation and exploration. We don’t really quite understand what this thing is that we’re building. I’m so glad you’re running this series, getting people to try to unpack it and figure out how we wrestle with it. There are clearly some very urgent questions around how to reduce harms and difficulties, but I think that the tooling that people are building, especially, is making it more and more accessible for people to come and play in this area and bring some of those different perspectives to try to understand what it is that we have, what these tools can do for us, and what the next decades are going to look like...” –

Amy Heineke (Academia 01:28:58)

Thanks to Deep Dive: AI, we are enlightened and much better informed, but we are still seeking answers. In the coming year, OSI will delve deeper into the conversations and collaboration that need to occur to support AI innovation and protect society and the proactive role we can take in that process. Stay tuned for our invitations to join in the effort.

Acknowledgements

DEEP DIVE: AI CONTRIBUTORS AND SPONSORS

The OSI extends its gratitude to the esteemed individuals who contributed their wisdom and experience to our discussions and to the corporate sponsors who made this event possible.

Contributors

Albon, Chris
Director of machine learning, Wikimedia Foundation
Twitter

Biderman, Stella
Mathematician and AI researcher, EleutherAI
stellabiderman.com, Twitter, Sigmoid.social

Carlberg, Astor Nummelin
Executive director, Open Forum Europe
LinkedIn, Twitter, Mastodon

Chestek, Pamela
Principal of Chestek Legal, and Chair of the OSI License Committee, OSI board director
LinkedIn, Fosstodon

Contractor, Danish
Chair, Responsible AI Licensing Initiative
LinkedIn, Twitter, Mastodon

Haddad, Ibrahim, Ph.D.,
General manager, LF AI & Data Foundation
LinkedIn, Twitter

Heineike, Amy
Vice president of engineering, 7bridges
LinkedIn, Twitter

Jalali, Adrin
Maintainer, Sckikit-learn Fairlearn, Hugging Face
LinkedIn, Kolektiva.social

Kanter, David
Executive director, MLCommons
LinkedIn, Twitter, Email

Kimmich, Sal
director of Open Source, AI DevSecOps, EscherCloud
LinkedIn

Leahy, Connor
ounder of EleutherAI and CEO at Conjecture
LinkedIn, Twitter

Lee, Jennifer
Technology and liberty project manager, ACLU-WA
LinkedIn, Twitter

Muñoz Ferrandis, Carlos
AI counsel, Hugging Face
LinkedIn, Twitter

Surman, Mark
Executive director, Mozilla Foundation
LinkedIn, Twitter

Tarkowski, Alek
Strategy director of Open Future Foundation, board member of Creative Commons
Indieweb, Twitter, Mastodon, LinkedIn

Villa, Luis
Co-founder and general counsel, Tidelift
Twitter, Mastodon

Walsh, Kat
General counsel, Creative Commons
LinkedIn, Twitter

Walsh, Kit
Senior staff attorney and assistant director, Electronic Frontier Foundation
Twitter

Widder, David Gray, Ph.D.
Candidate, Carnegie Mellon University School of Computer Science
Website, Twitter, Mastodon

Zhou, Mo, Debian Developer, Ph.D.
Sudent at Johns Hopkins University
Website, LinkedIn, Mo Zhou at Debian.org

Sponsors

No sponsor had any right or opportunity to approve or disapprove the content of this event.

GitHub DATASTAX Google

“

AI holds promise to expand developer capabilities and opportunities in multiple dimensions.

It allows more developers to use advanced tools like formal methods, lowers barriers to becoming a developer through accelerated learning, and increases software quality while decreasing its cost, which will increase the overall opportunity and demand for developers, as open source has done for decades.

Mike Linksvayer

Head of Developer Policy
GitHub