Datasaur

# Guide to Liability in AI

November 2022 | Guides

# 01 Introduction

We're living through a technological revolution. And, while this is exciting, it also means we're walking through uncharted territory. We're having to answer questions and tackle ethical issues that we could never have imagined.

As the technologies develop, they are often doing so before we have a chance to completely consider—and mitigate—the ethical impact that they will have. Take the advent of AI-generated art tools like Stable Diffusion, Midjourney, and DALL·E as an example. These technologies are surfacing questions around ownership, regulation, and the potential destructive impact of art in ways that we've never seen before.

Now that generative AI tools exist, we're having to consider whether to ban political or sexual content in digital art. We're having to consider the ways that peoples' digital rights could be infringed by deepfakes. And this is just the tip of the iceberg.

Not only do we have to consider these ethical questions, but we also have to see them play out with new angles, and at a scale that we've never seen before.

And in many ways, our current liability system simply isn't built for the technologies and the impact they're having. This means that we're having to adjust and evolve as technology evolves.

# 02 How to Assess the Ethical concerns for AI

The question of where to assign liability with AI is complex, layered, and evolving. There is also no cut-and-dry answer.

With that in mind, this guide will look at ethics and liability within three specific use cases of AI and NLP:

1. Household robots
2. Self-driving cars
3. AI-generated content

We will assess each of the use cases by considering two factors that pertain to liability: the amount of potential harm/impact, and agency.

# 03 How Use Cases of AI Underpin Ethical Questions

When people talk about AI and ethics, there's a tendency to talk about AI in its entirety. But this runs the risk of being reductionist since the technologies and use cases of AI are so varied.

Perhaps, for this reason, use cases are the backbone of the EU's proposed AI Act. The Act establishes different requirements and regulations for AI in high-risk use cases—such as those that affect a person's health—compared to lower-risk use cases, such as marketing. Companies using AI for high-risk uses will be required to conduct more rigorous testing of the software and must maintain a strong audit trail.

Let's look at our three use cases of AI: **household robots, self-driving vehicles, and AI-generated content**. These are wildly different technologies and there isn't a one-size-fits-all answer for where liability should lie that applies to all three. It's imperative to look at them individually.

## How do we Assign Liability with AI?

A couple of factors that decision-makers are turning to when drawing up legal frameworks for each of AI's use cases are:

1. **Amount of potential impact/harm**
   It goes without saying that there is a big difference between the amount of harm that a self-driving vehicle can inflict compared to a floor-cleaning robot. The level of potential harm helps to dictate how the technology should be regulated.
2. **Agency**
   In the US, agency is a key condition for legal responsibility. Here, agency is broadly defined as having the ability to act and to do so with intentionality. This can inform us with AI as well. For example, with AI content generators, the tool itself doesn't have agency. Rather, it is fully programmed by humans to follow human-led rules. Therefore, it follows that liability and responsibility should likely be placed either on the humans developing the tool, or on the users generating the content.

Amount of potential harm and agency are key parts of decision-making around AI. For example, the EU's AI Act proposes that when harm can be established and an AI system was involved in the decision-making (and there is some likelihood that the AI software contributed to that harm) there will be a presumption of liability.

This means that anyone developing AI technology should be aware of the level of agency, potential harm, and use cases.

# 04 Household Robots

## How Advanced Are They?

Household robots are making great strides. In fact, two tech giants are hoping to launch household robots in 2022: Tesla is developing the Tesla Bot and Google is developing PaLM-SayCan. Both bots are being designed to complete household tasks like cleaning and making dumplings.

Arguably the most sophisticated bipedal robot is created by Boston Dynamics. It's complex and has taken decades to develop.

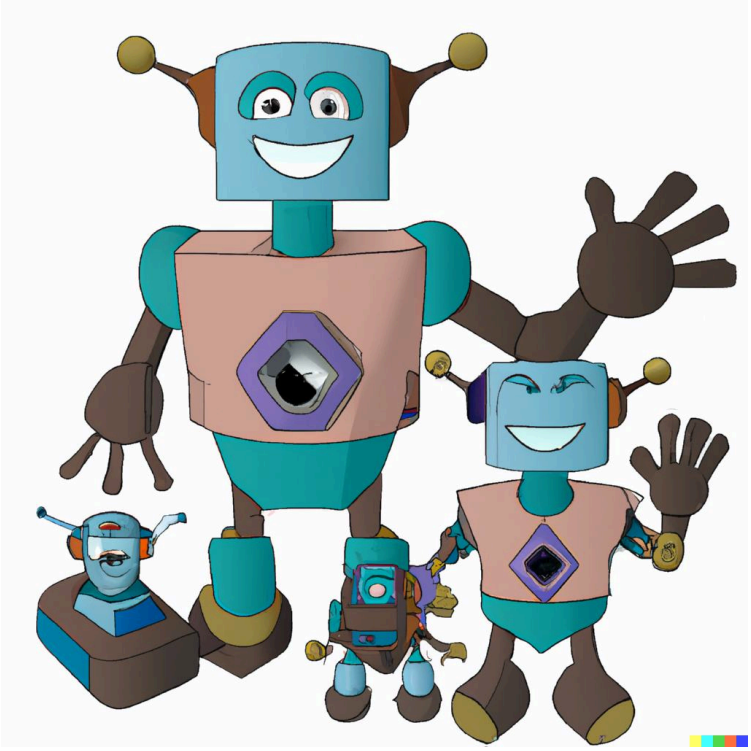

*Fig. 1: AI-generated art from DALL-E*

## How Much Liability Exists Around Household Robots?

Currently, none of the bots (Tesla's, Google's, or Boston Dynamics') are reliable or high-performance enough to be sustained in a real-world environment. Ultimately, the tech community largely believes that household bipedal robots could be achieved within a decade but we are not there yet.

This means that right now, their amount of potential harm is negligible.
It also means that rather than looking at liability for damage right now, we are at the point where we're assigning liability and responsibility for future applications, and are safeguarding against potential damage when the robots are ready for real-world use in peoples' homes.

What's more, household bots aren't deemed to have agency or the capacity to pass the Turing Test. Debates have sprung up recently about whether robots and bots will gain sentience and agency in the future, but we certainly aren't there yet. This means that liability for potential damage must fall on the manufacturers or users.

## So Where Does Liability Lie for Potential Damage?

Tech companies are already assuming a high level of responsibility, and are putting restrictions in place to limit the robots' power. For example, Tesla's bot is being limited in size and restricted to running 5 miles per hour. Their explicit intention is for humans to easily overpower it.

As the robots get closer to being ready for real-world application, we will have to decide whether liability for potential damage falls on the user, company, or some combination of the two.

We don't have the answers yet, but we do know that AI companies are already considering regulations, restrictions, and liability—and the world is watching as they do.

# 05 Self-Driving Cars

## How Much Liability Exists Around Self-Driving Cars?

Self-driving cars carry a high level of potential harm, and the accidents they create have attracted a lot of press attention. This has created a huge amount of pressure around assigning liability.

As with household robots, the AI system itself is not deemed to have any agency or sentience, so liability is falling either on the manufacturer or the driver. Deciding between the two has been a source of great—and very public—debate.

### Where Does Liability Fall for Accidents?

The majority of laws and clauses around self-driving cars currently put responsibility and liability onto car manufacturers. In fact, the EU recently announced that manufacturers of self-driving vehicles will be liable for a vehicle's actions when in autonomous mode.

However, many say that the current liability system we have is unprepared for AI. For example, most current liability systems still place some degree of liability on the human causing the accident, but there are cases with vehicles now where the human had no input at all.

There are also questions around what happens when the driver uses the AI system correctly but there is an unforeseen situation that the car's software isn't prepared for. For example, a self-driving car was put in Autopilot mode and struck and killed a pedestrian in that mode.

This illustrates the nuances of liability and how important these discussions—and decisions—are. Self-driving cars carry high levels of potential harm, and the AI cannot be deemed to have agency, thus the legal responsibility has fallen onto the people building and using the technology.

As we gain more data around self-driving vehicles and incidents, we will continue to regulate and assign liability and make adjustments as needed.

## 06 AI-Generated Content

AI-generated art is hitting the mainstream now, and more and more people are learning about tools like Midjourney, DALL·E, and Stable Diffusion. These open-source tools scrape images from the internet, and they have unleashed a slew of questions about ethics, ownership, and more.

It's a complex and evolving topic, so let's first take a look at it in terms of our two factors: the amount of potential harm and agency.

### AI-Generated Content and Potential Harm

Harmful content will always exist. And deciding what is harmful, who is liable for that content, and how to regulate it is a source of great debate.

How do we determine harm, and how do we protect people from that harm without infringing on freedom of speech and expression?

AI art generators have pretty stringent regulations in place already to prevent harm and mitigate their liability. Each of the major content generators does not allow for graphic violence or pornography, for example. This has been enforced with overwhelming success and should be praised for the fact that shocking/destructive artwork has not really come out.

There are more ways that harmful content can be created, though. For example, people can create lewd and explicit images now using a person's likeness. This opens the door to a great deal of potential harm for those people whose images are used.

We don't have the answers here yet, but we do know that AI art generators already have a huge number of users, and an even higher number of people who have been exposed to AI-generated art, so the potential impact is huge.

## AI-Generated Content and Agency

In terms of agency and AI art generators, the AI company (generally) isn't looking at the art before it goes out. It's automated and there's no human in the loop. And the AI tool itself does not have a sense of agency.

This means that the AI tool cannot be held liable, so liability must fall on the company or on the user. Since there's no human in the loop, the AI company cannot be held completely liable for the art, either. So it naturally follows that some liability must be assigned to the user/creator.

As a side note, we might imagine a future where AI content generators are deemed to have agency and sentience. What will it mean for any work that is created then? If we reach that point (and many NLP researchers believe we are inching that way) it will be fascinating to watch the discussions unfold.

## Where Does Liability Lie?

For now, AI companies have been carrying the bulk of liability by looking at their processes and regulations to make sure they're doing what they can to mitigate as much harm as they can. However, as discussed above, there are cases where the user will assume liability.

This is where the decision-making gets more complex. With AI-generated art and content, it can be difficult to assign ownership. AI-generated art is created by a user but is trained on huge numbers of artists' work.

How do we regulate this? Should all artists' images online be explicitly labeled by copyright intent? And is there a future where we can understand which images/artists were referenced and add that as a training note? This is non-trivial, because this is a huge undertaking, especially when the tools use so many images to come up with content, raising questions about how we assign influence proportionally and fairly.

For now, the bulk of liability has fallen on the AI companies, and some liability may fall on the user creating the content as well. It will be interesting to watch the way that this plays out as generative AI continues to evolve and expand its influence.

## Could Regulation Go Too Far for AI-generated Art?

*Fig. 2: AI-generated art from DALL-E*

One final thing to consider is how to regulate AI and hold the right parties accountable, without going too far and running into issues like censorship. There's a sweet spot between protecting freedom of speech and safeguarding society from harm from technology. And that sweet spot is hard to find.

Here are some concerns that are being raised right now:

1. **Generative AI content restrictions**

   AI art generators have content moderation in place that prevent over-sexualized or violent art. This is a good thing to protect users, however, there is some debate around the fact that art has been an outlet for expression since the beginning of time.

   People have used art to express injustice, opinion, imagination, political strife, and more. How do we find the line between freedom of expression and protecting society from harm?

2. **Societal norms**

   There is also debate around the fact that different societies have different accepted norms. For example, different cultures around the world have different accepted levels of sexual freedom. Marvel's Eternals movie was recently banned in Saudi Arabia as a result of a same-sex kiss, and other countries pushed for the cut to be edited.

   Much AI and NLP development so far has been Western-centric, so how does this apply to AI-generated art, and where is the line? This gets complicated very quickly. It is a variable that will have to be considered as regulations are created, as there are different precedents in different countries.

   For example, China's text-to-image generation AI systems are being built with restrictions that line up with the country's political censorship rules. The system does not allow requests for images depicting certain Chinese political leaders and images that the government considers politically sensitive.

3. **Loopholes and alternatives**

   Another issue is that regulating content does not fully prevent that type of content from being created. Humans are agile and creative and will find an alternative. This doesn't mean that we should not regulate the content; rather it is raised as a factor to consider when creating restrictions and regulations.

We don't have concrete solutions. We're only on DALL·E 2, after all. But restrictions are being built to safeguard the target audience. As more use cases come up we will need to iterate on the safeguards we have.

# 07 Conclusion

As AI develops, it's raising hard-hitting questions that we don't all have the answers to. Yet.

The more technology evolves, the more the conversation evolves with it. Great strides are already being made in regulating AI and its impact—as with self-driving vehicles—and that will continue to happen as more use cases come up.

For now, this is a critical conversation to keep an eye on and to watch unfold. It's mind-blowing to think about how different our laws, societies, and daily lives will be a few years down the road as AI continues to expand its impact.

If you'd like to learn more about Datasaur's stance, our goals, or how we can support your NLP needs, please reach out and we'd be happy to talk.

## About Datasaur

Datasaur is a private LLM provider and data labeling platform designed for companies to build their AI ecosystem with ease and efficiency. It assists organizations and universities in setting up custom LLMs and annotating data more efficiently and accurately through automation, quality control, and human-in-the-loop workflows. For more information, visit www.datasaur.ai.

Schedule a demo