AI tools and technologies are advancing at such a pace that it is almost dizzying to try to keep up. Every single frontier that utilizes AI tools, from security, to [BLANK], to even [BLANK] seems to constantly be upturned and evolving, with better tools that hardly seemed possible before. Of these, I think some examples are not only important through the progress they represent, but also how they will affect companies, better our lives, and protect our future.

One of the first tools that caught my eye was Amazon's Bedock Guardrails, which prides itself on being a tool that can be used by developers to ensure that their models not only reach a higher level of efficiency, but efficacy as well. It does many things from checking factuality/preventing hallucinations to keeping models correct and on topic. It lets developers define off-limit topics to filter undesirable test strings and images (even for things like financial investment advice), enables barriers to ensure that personal identifiable information (PII) doesn’t leak into inputs and outputs, and even allows for developers to customize these settings to a wide range of models. This contribution better allows for developers to custom tailor their models to ensure that people still get the use of the model whilst protecting company, personal, and sensitive information whilst also better keeping rogue models in check. If practices like these become more mainstream, it will affect basically any mass distributed model and benefit everyone. As long as technology like this keeps on advancing and better maintaining models, we only benefit in my opinion.

Another company I found interesting was KPMG, which focuses on looking into the processes of specific companies and then designing AI tools to increase their efficiency. The base focus of the company is auditing other companies and then working alongside their teams to develop and enable better business practices. In the pursuit of this, they have recently begun utilizing AI tools to revolutionize how they assist companies and redefine what an audit company does. If this continues to progress as it is, this will better allow us to transition into a more automated and more cost effective society allowing more time to work on the things that matter.

One huge change recently is probably our biggest next step to the ultimate goal at the moment: Artificial General Intelligence (AGI). The proponent that made the advancement this time is OpenAI’s new model o1. The main difference between o1 and previous Large Language Models (LLM’s) is its ability to understand a prompt and reason a response, rather than simply respond with tokens that best match the prompt given. In other words, while models were already pretty good at summarizing or translating texts, they often fell short when matters got more complex. But with this model, it marks the first step in models that analyze the text, taking a step back and reasoning multiple reasonable responses, and deducting the response that best answers the prompt. This is huge and a major change and advancement that very well could be the next chain link that connects us to AGI.

Starting with Australia, they seem to want to give suggestions/guidelines that act as road marking for companies to allow them to better create safe, ethical deployments of AI models. In their ten voluntary safety standards, they emphasize the necessity of considering oversight, transparency, and full accountability throughout the entire AI development pipeline. To achieve these ideals, they have three main sections: Mandatory High Risk AI guardrails, National AI Ethic Principals, and Sector to Sector Regulations. For the first one, it gives the government proper authority to have more insight and control over the development of AI’s which could serve a greater risk to anyone. This would allow them to better test/audit the companies developing these AI, force better transparency, and easily define what is exactly expected of these companies if they choose to pursue such an endeavor. Next, Australia has outlined eight national AI ethic principles that set the stage for what should be developed and how. These range from discussing fairness/accessibility/usability (no discrimination), privacy protection, robust/safe models, and others. Lastly, Australia expects for the use of models to be used properly for things like finance and education, as misuse or abuse could easily lead to issues. In general, Australia is working alongside others to attempt to leash AI development and ensure that its progress is a benefit to everyone.

The European Union (EU), similar to Australia, has officially outlined their plans to regulate AI development through their own AI Act, which also emphasizes the need to promote models that maintain peoples rights, privacy/safety, and basic ethics. In this act, they have a classification system with four levels of risk associated with them, Minimal-no risk (video games, spam filters, etc.), limited risk (systems like chatbots must be transparent), high risk (healthcare, transportation, police, subject to stringent mitigation outlines), and unacceptable risk (threat to public safety and fundamental rights). Based on these, whilst the lower two have very few regulations, high risk developments must provide plans to show how they minimize risk, ensure high quality data and processes, maintain up to date quality documentation, have human oversight, and abundantly ensure robust, secure systems. This act also creates a new authority in EU supervisory authorities who, sourced from every EU state, will act as facilitators of these new regulations. If companies do not comply with these rules, they will be looked at by these authorities and subject to penalties that can range up to 30 million euros.

Overall, whilst both clearly take the issue of AI seriously, I find the EU’s system to be more robust, reproducible, and effective. Technically they do the same thing, with low risk AI’s mostly being ignored whilst high risk and heavily regulated, but it seems that the EU has a far more structured system for determining what is high risk, and what is not, whilst also creating a new authority uniquely positioned to follow through with its promises. On the other hand, if there were AI superhero vigilantes, Australias would definitely have a better system as stringent laws only serve to restrict the movements of justice seekers like Batman. Though, the EU definitely would be more likely to find out if my Amazon Alexa fridge has been recording all of my thoughts via electric waves or some other phenomena. While I am not a hero/villain expert, this is my best analysis (lol).

In order to look to the future, I believe that it is imperative to identify the trends of the past and how that trajectory shoots through the current frame of how we see the world. AI has only recently become a major factor in everyday life, but in this short time, we have seen growth so fast that it's hard to fully comprehend what it might mean or how it will fully affect us. In a way, we are in the future, but still think in the same way we did in a past age (imagine old farmers moving to the city in the middle of the industrial revolution), so how can we adapt how we think to better imagine our future? I think seeing how AI is already used is a great example. At the moment, while still proprietary, models are already capable of completing basic tasks, 24/7/365. This improvement over humans is what allowed them to take over these jobs. And while we haven’t reached it yet, we are still trying to find places where AI will be more effective than their human counterparts, and I see this as the main drive for AI: company executives see AI as a cost cutting feature that will save them billions if implemented well.

Now, looking into future #1, while AI seems practical and all-knowing, we will soon find limitations in its ability, and will eventually learn to coexist. In a way, this relationship would reflect how we humans do complicated mathematics, whilst using a calculator (computer) to handle the repetitive and mundane stuff. This future entails a world of dreams and promises, where AI serves as a sort of motor that drives us wherever we choose to go, and while some jobs will disappear, new jobs will replace them as we better specialize our society to this new and advanced world. Maybe instead of a lot of factory workers, there will only be two or three people maintaining each factory. This would free up all the people who used to work there who could now work in other fields such as the sciences, engineering, healthcare, or even data generation/science, creating data for model training (like driving cars, mocap movement, etc.) While satisfying though, this future is very idealistic and I believe leaves out one component that drives how we function today: upper management.

In an alternate future, AI can easily be seen as a tool of oppression, not necessarily in an authoritarian way, but in a capitalist-oligarchic way. If we continue to give more power to the top percentiles, whilst also giving them the ability to replace their need for us, the average person can quickly be seen as an unnecessary tool in their world. At the moment, they need us to man their banks, engineering departments, and other facilities, but what if they see AI’s capacity to replace us as another cost cut, and decide that if they suppress us, they also never have to worry about strikes, complaints, lawsuits, or even laws in general? The top percent already have an unfathomable amount of power in our current society and they very well may see AI as the last step to automating their lives whilst maintaining power.

With this in mind, I see regulation as proposed by these bodies of government as necessary and vital if we want to maintain our lifestyle and level of liberty, as we can easily lose it as soon, or even faster than we gained it. And while AI can serve for good and create limitless possibilities for everyone in society, it is merely a tool and can just as well be weaponized as a replacement for the need for a general populace, and render us useless in the eyes of those in the top percent.

References

<https://www.youtube.com/watch?v=hACwBNRR3to&t=138s>

<https://aws.amazon.com/blogs/machine-learning/advancing-ai-trust-with-new-responsible-ai-tools-capabilities-and-resources/>

<https://aws.amazon.com/bedrock/guardrails/>

<https://kpmg.com/us/en/capabilities-services/audit-services.html?utm_source=chatgpt.com>

<https://www.reuters.com/technology/artificial-intelligence/kpmg-spend-100-million-ai-partnership-with-google-cloud-2024-11-20/?utm_source=chatgpt.com>

<https://www.theatlantic.com/technology/archive/2024/12/openai-o1-reasoning-models/680906/?utm_source=chatgpt.com>

<https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-principles/australias-ai-ethics-principles?utm_source=chatgpt.com>

<https://www.reuters.com/technology/artificial-intelligence/australia-plans-ai-rules-human-oversight-transparency-2024-09-05/?utm_source=chatgpt.com>

<https://www.reuters.com/technology/nvidias-business-practices-eu-antitrust-spotlight-sources-say-2024-12-06/?utm_source=chatgpt.com>

