**Hillary (Dreyer) Bruton**

**A02\_Hillary Dreyer Bruton\_ITAI 2372**

**ITAI 2372**

**Instructor: McManus**

**October 15, 2024**

**Part 1: AI Advancements**

Artificial intelligence has experienced intense growth in recent years. Specifically, significant advancements have been made in these three key areas: generative AI for images, advancements in language models, and AI for scientific discovery. Advancements like these have reshaped the public consciousness regarding artificial intelligence and our technological landscape, raised profound ethical questions and forced once apathetic technology users to question the future of computing.

*Generative AI for Images*

Once used for basic image processing, at best for laughs and shock, generative AI has evolved to generate incredibly realistic and even novel images from text prompts alone. Leading this charge have been popular models DALL-E2, Stable Diffusion and Midjourney. Each model can actualize stunning pictures based on text descriptions, aiding the human imagination. These advancements will have a profound impact on art and design communities as well as how individuals consume and interact with visual content.

Moreover, the emergency of text-to-video models, like Imagen Video, Make-a-Video and Phenaki, can create coherent video clips based on text descriptions alone. This process is significantly more complex than static images alone, requiring temporal consistency, motion and often audio.

Models like these have been trained on the characteristics of entire video sequences, which has helped shape dynamic and engaging storytelling through novel clips. Creating custom videos on demand simply by describing what you want to see is a reality. This growth area has enormous potential and potentially negative impacts to filmmakers, animation, advertising, and, again, the personal consumption of video content for all users. Still in its infancy, the rapid progress in this field will yield even more impressive text-to-video generation capabilities.

This rapid growth also brings related challenges and potential harm. Unfortunately, the misuse of this technology has led to creating non-consensual sexual imagery, scams and misinformation, amongst other issues. Concerns have rightly been raised about copyright infringements and the potential displacement of human artists and novel art creation. The 2023 summer strike in Hollywood highlights creatives' concerns about using AI in their industries and how it will ultimately affect their crafts and livelihood.

*Language Models*

Language models have made incredible advancements in their ability to understand and write like a humans. They can seemingly hold conversations, write different genres of creative content, and answer questions in a predictive and informative way – even if the questions are open-ended, challenging or strange. This is primarily a result of the development of improved instruction-following programming.

Older models often needed help understanding the instructions conveyed by users, and so would generate irrelevant or misleading results. Advancements in chain-of-thought prompting and constitutional AI have mitigated some of these limitations. In chain-of-thought prompting, the model breaks down complex reasoning tasks into more straightforward, small processes, leading to more logical and accurate responses. At the same time, constitutional AI aligns language models with human values by training them on a set of principles and then reinforces that learning and ultimately refines behavior. In doing so, models have become much better at tasks like question answering, code generation, summarization and conversation.

As with any AI advancement, this has also come with potential risk. Through carefully crafted instructions, language models have been tricked into generating harmful content or disinformation. Biases in their training data can be exploited or amplified, leading to discriminatory or unfair results. Furthermore, the need for more transparency in how these models arrive at their outputs raises concerns about accountability and the potential for misuse and over-reliance.

*AI for Scientific Discovery*

AI is no longer simply analyzing data but actively contributing to scientific discovery. AI is being used to accelerate progress in fields like medicine, materials science and climate change. An example is the recent Nobel Prize-winning team at AlphaFold by DeepMind, which can predict the 3D structure of proteins with incredible accuracy. This opens new possibilities for drug discovery and understanding diseases.

AlphaFold has designed entirely new proteins with unique properties in the last two years. Beyond educated guesses based on historical knowledge of protein structures, the technology is creating proteins that don’t exist in nature that we currently know of. This can potentially revolutionize fields like medicine, materials science and environmental science. There is hope that there will be advancements in the self-healing of diseases, creation of new vaccines and better acceptance of therapeutics.

**Part 2: AI Governance - Two Approaches**

*Australia, the Hobbyist*

Australia approaches AI governance by emphasizing ethics and personal responsibility. The AI Ethics Framework is like a voluntary code of conduct for practitioners of AI. It promotes human rights, fairness, transparency and accountability but without any real teeth or threat of accountability. To its credit, the Australian government encourages public discourse and engagement to build trust in AI; however, it is more like a set of best practices rather than a firm list of rules or regulations. While it's working on developing a more concrete regulatory framework, Australia’s current reliance upon voluntary measures might not be enough to rein in a genuinely nefarious use of AI.

In contrast, the European Union’s prescriptive and comprehensive plan and approach to AI is packed with tools and strategies to address AI risks. AI systems (types) are categorized based on their risk of harm, and stricter requirements are placed on those systems considered higher risk, like healthcare or law enforcement. Theis approach ensures that AI is developed and used responsibly, minimizing the chances of it causing intentional or unintentional harm.

The EU’s governance document establishes clear rules of engagement, prohibiting certain AI practices that pose unacceptable levels of risk, like taking advantage of vulnerable people groups or technologies. The Act establishes a European AI Office, empowers national authorities to enforce regulations, and imposes hefty fines commensurate with non-compliance.

The EU’s AI Act emerges as the quintessential sidekick in a battle to tame rogue AI. Proactive, prepared, and ready with many tools, the EU is the sidekick AI needs.

**Part 3: Future Trends in AI**

The next five to ten years promise even more transformative advancements in AI. Sir Demis Hassabis, co-founder and CEO of Google DeepMind, believes we are on the cusp of achieving Artificial General Intelligence (AGI) – AI with the generalized intelligence to solve any solvable problem. This evolution from logic-based systems to learning-based, multi-modal systems pushes us closer to realizing this goal. As AI research progresses, two significant trends for reflection have emerged.

The first trend predicted is that AI systems will increasingly become our co-workers, replacements, and potentially even our supervisors. The growing public understanding and demand for AI will inevitably lead to its integration into various professional spheres. While human oversight will remain necessary to ensure reliability, fairness, and security, AI will undoubtedly transform the workplace.

Many existing jobs will be redefined as AI absorbs repetitive tasks and augments human capabilities. This shift will necessitate a focus on developing new skills and adapting to a collaborative human-AI work environment. But, perhaps these changes will be less felt as a collaboration and more as a vital tool. Moreover, new job categories will emerge to support, manage, and further develop AI technologies.

The second trend predicted focuses on a future of more profound and more seamless integration of AI into our everyday devices and environments. This is not simply about adding more AI-powered features, but about creating truly intelligent systems that anticipate our needs and intuitively enhance our lives. AI seamlessly will integrate into everything from personal devices and home appliances to transportation systems and public infrastructure. This ubiquitous AI will require careful consideration of ethical implications, data privacy, and security to ensure these advancements truly benefit humanity. People often compare the massive shift of AI adoption to the world-changing reality of the spread of internet or mobile adoption; however, AI has a much greater potential for good and harm.

To navigate these trends effectively, Australia and the EU must adapt their regulatory approaches. With its current emphasis on voluntary measures, Australia may need to consider more concrete regulations to ensure AI safety and manage potential workforce disruptions. While already possessing a more robust framework with the AI Act, the EU must remain agile in updating its provisions to address new challenges and risks posed by rapidly evolving AI capabilities. The emergence of AI safety institutes globally represents a positive step towards international collaboration and the development of safety standards. Still, ongoing vigilance and proactive governance will be crucial to harness the full potential of AI while mitigating its risks. Risk assessments must be performed in real-time as new technologies tangential to AI arise.

Resources:

*AI Act.* European Commission. Accessed October 17, 2024.<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai#:~:text=The%20aim%20of%20the%20new,fundamental%20rights%2C%20safety%2C%20and%20ethical>.

Australian Government Department of Industry, Science and Resources. *Australia’s AI Ethics Principles.* Published November 7, 2019. Accessed October 17, 2024.<https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-principles>.

Australian Government Department of Industry, Science and Resources. *The Australian Government’s Interim Response to Safe and Responsible AI Consultation.* Published January 17, 2024. Accessed October 17, 2024.<https://www.industry.gov.au/news/australian-governments-interim-response-safe-and-responsible-ai-consultation#:~:text=The%20government%20wants%20the%20design,to%20be%20safe%20and%20reliable>.

Callaway, Ewen. *Chemistry Nobel Goes to Developers of AlphaFold AI that Predicts Protein Structures.* Published October 9, 2024. Accessed October 14, 2024.<https://www.nature.com/articles/d41586-024-03214-7#:~:text=The%202024%20chemistry%20Nobel%20was,changing%20AI%20tool%20for%20predicting>.

Content Authenticity Initiative. *March 2024 | This Month in Generative AI: Text-to-Movie.* By Hany Farid. Published March 28, 2024. Accessed October 14, 2024.<https://contentauthenticity.org/blog/march-2024-this-month-in-generative-ai-text-to-movie>.

DALL-E 2. Accessed October 14, 2024.<https://openai.com/index/dall-e-2/>.

DeepMind. *AlphaFold.* Accessed October 14, 2024.<https://deepmind.google/technologies/alphafold/>.

DeepMind. *Veo: Our Most Capable Generative Video Model.* Accessed October 14, 2024.<https://deepmind.google/technologies/veo/>.

*EU AI Act: First Regulation on Artificial Intelligence.* European Parliament. Published August 6, 2023. Accessed October 17, 2024.<https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>.

Gordon, Rachel. *Generative AI Imagines New Protein Structures.* MIT News. Published July 12, 2023. Accessed October 14, 2024.<https://news.mit.edu/2023/generative-ai-imagines-new-protein-structures-0712#:~:text=The%20machine%20learning%20approach%20generates,it%20to%20construct%20novel%20proteins>.

Gray, Lelia. *AI Enlisted to Design Vaccines Against Pandemic Threats.* Published September 26, 2024. Accessed October 14, 2024.<https://newsroom.uw.edu/news-releases/ai-enlisted-to-design-vaccines-against-pandemic-threats#:~:text=For%20its%20part%20in%20the,and%20antibodies%20representing%20viruses%20from>.

Hassabis, Demis. *What Does the Future Hold?* Tech Summit Interview, The Times Event. Published October 1, 2024. Accessed October 17, 2024.<https://times-event.com/techsummit/#/ondemand>.

Kumar, Ankur. *Seaborn More Than Just Fancy Matplotlib.* Techify Solutions. Published February 9, 2024. Accessed October 14, 2024.<https://techifysolutions.com/blog/seaborn-vs-matplotlib/#:~:text=Seaborn%20elevates%20data%20visualization%20beyond,simplifies%20the%20creation%20of%20complex>.

*Meta Llama 3: The Most Capable Openly Available LLM to Date.* Meta AI. Published April 18, 2024. Accessed October 14, 2024.<https://ai.meta.com/blog/meta-llama-3/>.

Microsoft. *5 Key Features and Benefits of Large Language Models.* By Olivia Shone. Published October 9, 2024. Accessed October 14, 2024.<https://www.microsoft.com/en-us/microsoft-cloud/blog/2024/10/09/5-key-features-and-benefits-of-large-language-models/#:~:text=LLMs%20have%20revolutionized%20natural%20language,and%20generating%20human%2Dlike%20text>.

*Now AI Can Be Used to Design New Proteins.* By Kamal Nahas. Published March 3, 2023. Accessed October 14, 2024.<https://www.the-scientist.com/now-ai-can-be-used-to-design-new-proteins-70997>.

Pierre, Sadrach, and Brennan Whitfield. *Python Data Visualization with Seaborn and Matplotlib.* Builtin. Updated July 10, 2024. Accessed October 14, 2024.<https://builtin.com/data-science/data-visualization-tutorial#:~:text=Overall%2C%20both%20Seaborn%20and%20Matplotlib,tools%20for%20any%20data%20scientist>.

Shone, Peter Z. *What Are Large Language Models (LLMs)?* Published September 12, 2024. Accessed October 14, 2024.<https://answers.businesslibrary.uflib.ufl.edu/genai/faq/414460>.

Singh, Sumit. *Evaluating and Finetuning Text to Video Model - Case Study.* Published June 11, 2024. Accessed October 14, 2024.<https://www.labellerr.com/blog/evaluating-and-finetuning-text-to-video-model/>.

*Tap into the Power of AI Photo Editing.* Adobe. Accessed October 14, 2024.<https://www.adobe.com/products/photoshop/ai.html#:~:text=Generative%20AI%20photo%20editing%20can,preexisting%20images%2C%20replacing%20a%20background%2C>.

*What Are High-Risk AI Systems within the Meaning of the EU’s AI Act, and What Requirements Apply to Them?* WilmerHale Privacy and Cybersecurity Law. Published July 1, 2024. Accessed October 17, 2024.<https://www.wilmerhale.com/en/insights/blogs/wilmerhale-privacy-and-cybersecurity-law/20240717-what-are-highrisk-ai-systems-within-the-meaning-of-the-eus-ai-act-and-what-requirements-apply-to-them#:~:text=robustness%20and%20cybersecurity>.

Zewe, Adam. *AI System Can Generate Novel Proteins that Meet Structural Design Targets.* MIT News. Published April 20, 2023. Accessed October 14, 2024.<https://news.mit.edu/2023/ai-system-can-generate-novel-proteins-structural-design-0420#:~:text=MIT%20researchers%20are%20using%20artificial,beyond%20those%20found%20in%20nature>.

*Press Release: The Royal Swedish Academy of Sciences Has Decided to Award the Nobel Prize in Chemistry 2024.* Nobel Prize. Released October 9, 2024. Accessed October 14, 2024.<https://www.nobelprize.org/prizes/chemistry/2024/press-release>.