Class 1: Introduction

DPI-681M
The Science and Implications of Generative AI
Profs. Goel, Levy, and Svoronos
Harvard Kennedy School

1. Overview
2. Case: GAI in University Admissions
3. Intro to the Science of GAI

I. Overview

I. Overview

DPI-681M

Intro to the course

Three big questions we aim to answer together:

- **Unit 1:** How does generative AI work?
- **Unit 2:** How can I use it effectively?
- **Unit 3:** How will it affect the people and issues I care about?

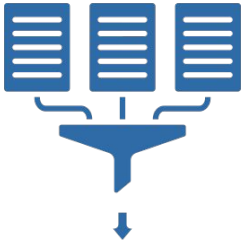
The debut of mass market LLMs

- ChatGPT released November 30, 2022
    - Not the first publicly available large language model (LLM), but heralded the beginning of widely used LLMs in non-technical hands

The debut of mass market LLMs

- ChatGPT released November 30, 2022
    - Not the first publicly available large language model (LLM), but heralded the beginning of widely used LLMs in non-technical hands
- Within 5 days it had 1 million users.
- On December 3, Aaron Levie, CEO of Box tweeted:

Why all the fuss? Why did you spend 125 points to get into this course?



Your uses of Generative AI (GAI)

1. **Creative & Artistic Uses** (e.g., generation of art, poetry, music)
2. **Academic & Research Assistance** (e.g., project brainstorming, citation generation, summarizing articles)
3. **Application in Coding & Data Visualization** (e.g., crafting Python code, creating graphs and unique visualizations)
4. **Drafting Documents & Content Generation** (e.g., drafting invitations, advocacy plans, internship applications)

How does GAI differ from "traditional" AI?

Prediction can be very useful, particularly when we need to **allocate limited resources**.

How does GAI differ from "traditional" AI?

Prediction can be very useful, particularly when we need to **allocate limited resources**.

- **Building inspectors** can prioritize buildings that have elevated fire risk.
- **Tax auditors** can prioritize individuals who have high likelihood of underpayment.
- **Hospitals** can provide transportation for patients at high risk of not appearing at their appointments.

What's new about Generative AI?

The latest generation of **generative artificial intelligence** can produce novel content — including **essays**, **computer code**, and **artwork** — that is often hard to distinguish from materials produced by humans.

Generative AI

According to a Goldman Sachs report, 3 key features of the technology relative to traditional AI that make it potentially impactful and quickly adopted:

Generative AI

According to a Goldman Sachs report, 3 key features of the technology relative to traditional AI that make it potentially impactful and quickly adopted:

1. **Generalized** rather than specialized use cases
2. **Generative** rather than descriptive - can produce original results indistinguishable from human output
3. **Approachable** interfaces that both understand and respond with natural language, images, audio and video

So how can we use it?

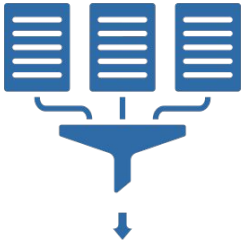
II. Generative AI in University Admissions

II. Generative AI in University Admissions

Why should you care about university admissions?

Focus on two questions

- ✓ How can **applicants** use generative AI?
- ✓ How can the **Admissions Office** use generative AI?



How can applicants use generative AI?

1. **Writing Assistance:** Essay draft, proofreading, and editing
2. **Interview and Test Prep:** Preparing for interviews (and standardized tests like GRE/GMAT/TOEFL
3. **Resume and CV Optimization:** Rewriting, including key experiences, listing achievements, and embedding values
4. **Research and Information Gathering:** Understanding program requirements, comparing programs, researching faculty and funding opportunities

🚴 Activity 🚴

- Get together with 1-2 classmates
- Pretend one of you is applying to HKS
- Write a 250-word essay using Gen AI
- Resources are here:

bit.ly/681-cpd

How good was the essay your team produced as a first draft?

A. Very bad

B. Bad

C. Neither good nor bad

D. Good

E. Very good

Focus on two questions

✓ How can **applicants** use generative AI?

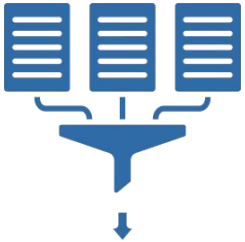
✓ How can the **Admissions Office** use generative AI?

Typical factors considered in admissions (in U.S.)

- Previous grades
- Admission test scores (SAT, ACT, GRE, etc.)
- Recommendation letters
- Essays
- Diversity

Typical factors considered in admissions (in U.S.)

- Previous grades
- Admission test scores (SAT, ACT, GRE, etc.)
- Recommendation letters
- **Essays**
- Diversity



How can the admissions office use generative AI?

1. **Screening and Filtering Applications** (Based on minimum requirements, GPA, resume checks, identifying AI-written essays, and initial screening)
2. **Enhancing Diversity and Fairness** (Assessing diversity and minimizing bias, ensuring a diverse cohort, and incorporating diversity in application processing)
3. **Summarization and Analysis of Essays and CVs** (Summarizing essays and resumes, detecting plagiarism, evaluating essay quality, and highlighting strengths)
4. **Improving Communication and Information Accessibility** (Creating chatbots for FAQs, drafting personalized emails, and producing promotional material)

How do we deal with “essays written by AI” problem?

🚴 Activity 🚴

- You will get 3 admission essays
- Assess if any of them was written (primarily or entirely) by ChatGPT
- Essay prompt: “Describe a time when interactions with others and/or an experience caused you to change your mind or expanded your point of view.”

* Adapted from and with great gratitude to Rosemary Hilliard (Senior Director of Admissions at HKS).

Written by ChatGPT?

A. Essay #1

B. Essay #2

C. Essay #3

D. None of them

Broader issues

- **Alignment problem:** Can HKS train generative AI tools well enough to advance its values in recruiting and admitting students? **-> Class 3**

Broader issues

- **Alignment problem:** Can HKS train generative AI tools well enough to advance its values in recruiting and admitting students? **-> Class 3**
- **Privacy**: If HKS trains its generative AI tool using applicants' essays and data, do applicants have the right to opt out? **-> Class 9**

Broader issues

- **Alignment problem:** Can HKS train generative AI tools well enough to advance its values in recruiting and admitting students? **-> Class 3**
- **Privacy**: If HKS trains its generative AI tool using applicants' essays and data, do applicants have the right to opt out? **-> Class 9**
- **Labor**: Will AI be labor-augmenting for HKS? **-> Class 10**

Broader issues

- **Alignment problem:** Can HKS train generative AI tools well enough to advance its values in recruiting and admitting students? **-> Class 3**
- **Privacy**: If HKS trains its generative AI tool using applicants' essays and data, do applicants have the right to opt out? **-> Class 9**
- **Labor**: Will AI be labor-augmenting for HKS? **-> Class 10**
- Other issues?

III. Intro to the Science of GAI

III. Intro to the Science of GAI

Large language models

“Large language models” [LLMs] are algorithms [like ChatGPT] that mimic human language.

At a high level, LLMs are built in three steps:

1. Collect a large **corpus of text** generated by humans;
2. Based on this text, **learn to predict** the next word in any given sequence of words;
3. **Fine-tune** the language model to **align** with desired behavior.

Step 1: Collect *a lot* of text

LLMs are typically trained on a massive dataset of web pages, online forums, books, and other content — comprising **trillions of words of text**.

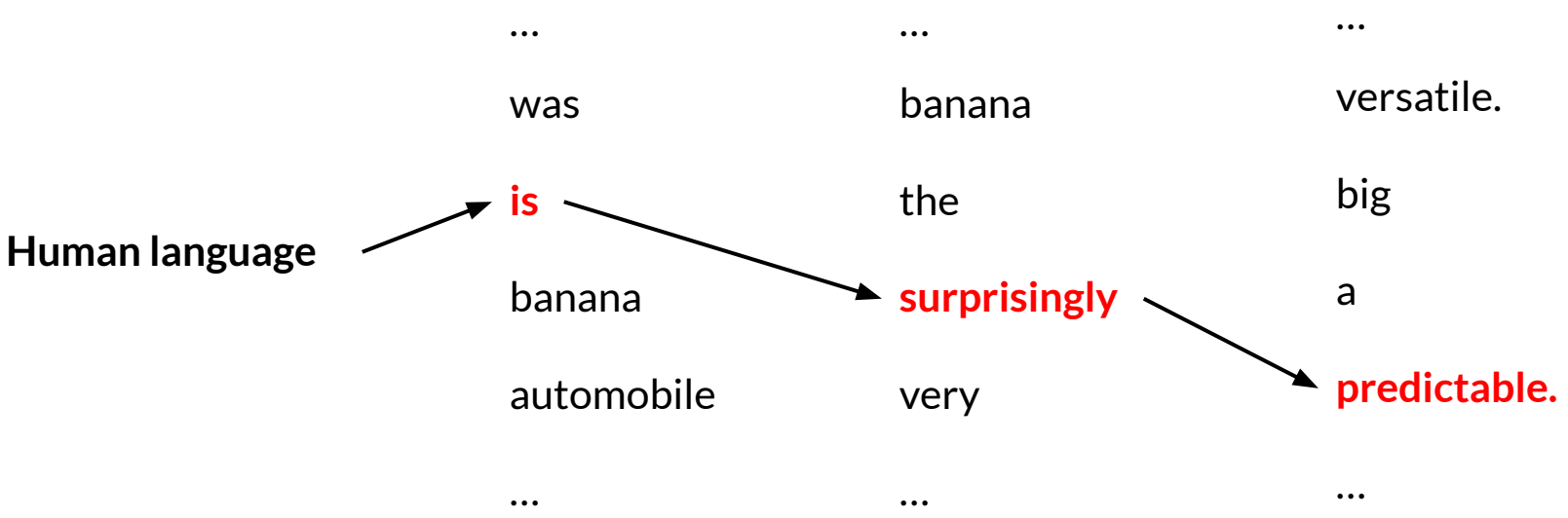
Step 2: Learn to predict words

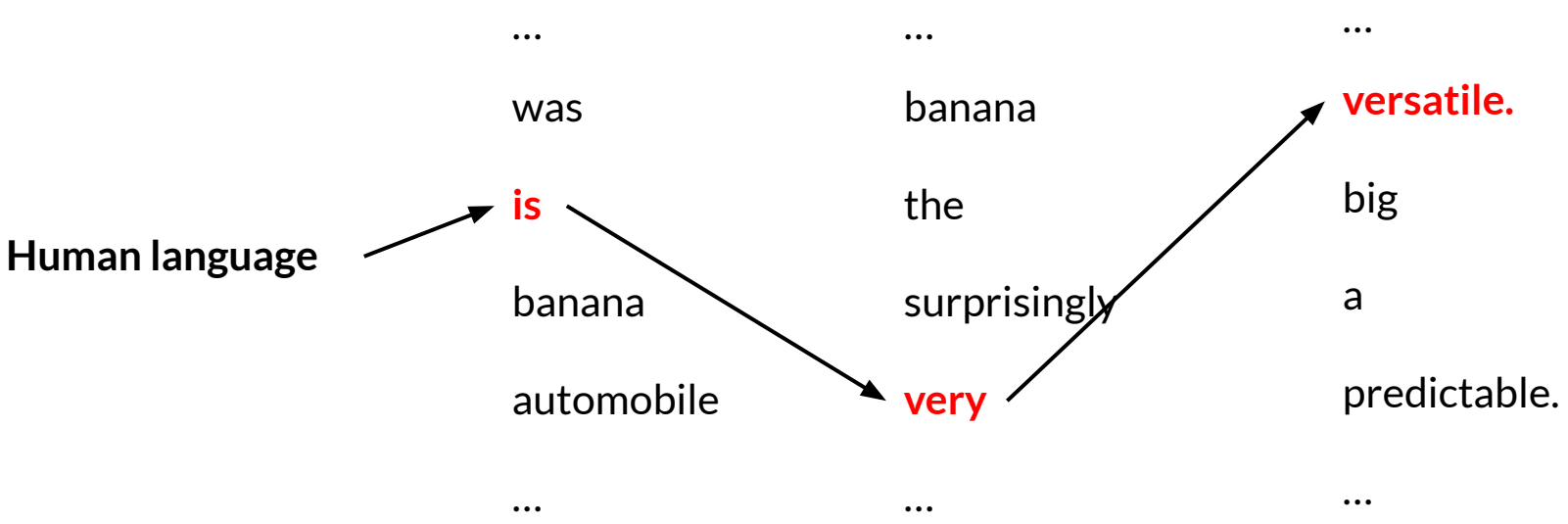
By analyzing this massive corpus of text, an LLM learns to **predict the next word** in a sequence of words. By repeating this process over and over again, it can generate long strings of coherent text.

Step 2: Learn to predict words

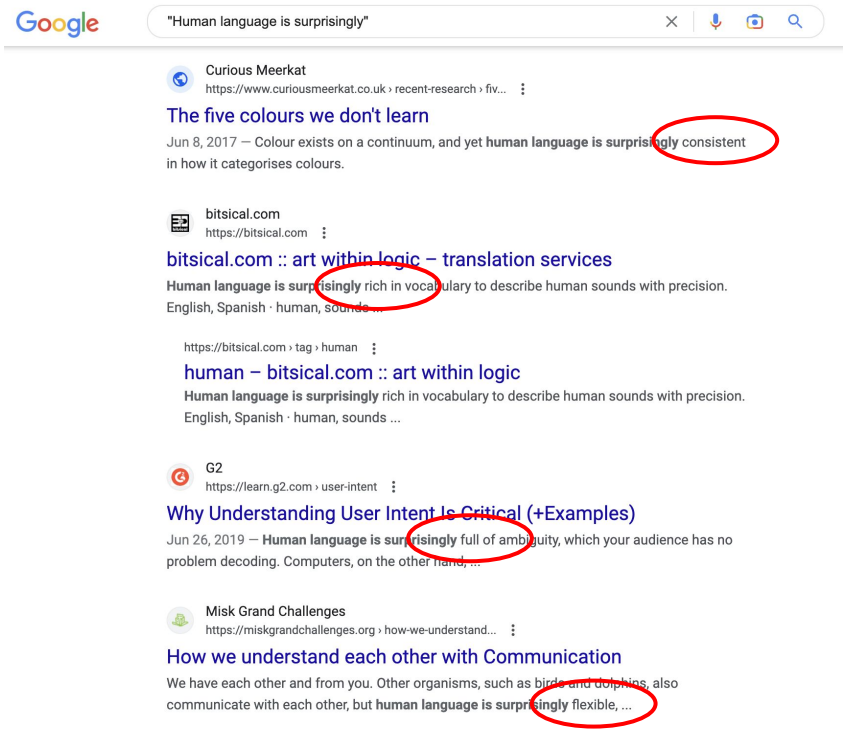
By analyzing this massive corpus of text, an LLM learns to **predict the next word** in a sequence of words. By repeating this process over and over again, it can generate long strings of coherent text.

To see this process in action, we'll collectively complete the following sentence one word at a time!
[Humans mimicking machines mimicking humans!]


Human language
...
was
is
banana
automobile
...
...
banana
the
surprisingly
very
...
...
versatile.
big
a
predictable.
...


Human language
...
was
is
banana
automobile
...
...
banana
the
surprisingly
very
...
...
versatile.
big
a
predictable.
...

Autocomplete on steroids



Autocomplete on steroids

Adult native-English speakers have a typical vocabulary size of about 30,000 words. The number of ways to string 10 of those words together is:

1,000,000,000,000,000,000,000,000,000,000,000,000,000,000,
000,000,000,000,000,000,000,000,000,000,000,000,000,000,
000,000,000,000,000,000,000,000,000,000,000,000,000!

[Of course, most of those combinations don't make sense, but there are still a **lot** of grammatically coherent possibilities!]

Autocomplete on steroids

Even relatively brief utterances are typically unique — they've never been said before and they'll never be said again. So we have no hope of finding exactly matching phrases in the data.

We instead need to build a statistical **model** of language.
[i.e., a large language model — more on this next time.]

Prediction is knowledge

Large language models can **encode knowledge** through **word prediction**, since correct answers typically involve **more probable** sequences of words than incorrect answers.

Prediction is knowledge

Imagine trying to predict the next word in the following sequence of words: **“The capital of Australia is _____”**

Prediction is knowledge

Imagine trying to predict the next word in the following sequence of words: **“The capital of Australia is _____”**

The next word is **more likely** to be “**Canberra**” than “**Sydney**” or “**Melbourne**”. [And much more likely than “**banana**”!]

Prediction is knowledge

To what extent can human knowledge, reasoning, and creativity be represented as a prediction problem that a language model can in theory emulate?

Appendix

Glossary

- **Artificial Intelligence**: a broad term that refers to any technology that is capable of intelligent behavior. This can include a wide range of technologies, from simple algorithms that can sort data, to more advanced systems that can mimic human-like thought processes.
- **Generative Artificial Intelligence**: a specific type of AI that is focused on generating new content, such as text, images, or music. These systems are trained on large datasets and use machine learning algorithms to generate new content that is similar to the training data. This can be useful in a variety of applications, such as creating art, music, or generating text for chatbots.

Source: https://www.antler.co/blog/generative-ai

Essay # 1 [Written by ChatGPT? __ Yes; __No]

Growing up in Casablanca, Morocco, where the hustle and bustle of the market was a male-dominated scene, I saw how our rich traditions co-existed with undervalued roles for women. This sparked my desire to promote female empowerment. At first, I thought change should come slowly to fit within the rhythms of my culture.

My time at the Paris School of Economics was a turning point. Paris was a different world, buzzing with new ideas. Here, I met Alice, a French classmate. We often debated over cups of coffee, discussing how best to achieve gender equality. I thought slow change would work best for Morocco. But Alice said we needed fast, big changes or risk making unfair treatment of women normal.

One winter evening, Alice invited me to a panel discussion on "The Urgency of Gender Equity" held at the Sorbonne. Listening to powerful stories and evidence from around the world showed the economic and social cost of gender inequality. That night, I realized respecting culture shouldn't stop progress or justice.

Now, my fight for gender equality has a new sense of urgency. I believe in the power of education and good policy to bring real change. I hope to learn more with a Master's in Public Policy from Harvard Kennedy School. I dream of the day when Moroccan women walk freely in the marketplaces, not just as customers but as successful shop owners. I'm committed to making sure no girl or woman should have to limit her dreams because she's a woman, no matter how hard it may be.

Essay # 2 [Written by ChatGPT? __ Yes; __No]

As a middle-class, suburban, American child of the digital age, I grew up with access to the latest and greatest technologies. For much of my life, I viewed technological advancement as the panacea for all socio-economic issues. My faith in technology was unshakable until a life-changing experience during a community outreach program in rural Yunnan, China while I was studying abroad.

I volunteered to teach basic digital skills to rural farmers, optimistically believing I could bridge the digital divide. The reality was startlingly different. The villagers, though polite and appreciative, seemed largely indifferent to my teachings. The smartphones and computers were alien to their traditional way of life and seemed to offer little tangible value to them.

Frustrated, I gradually fostered deeper conversations with them. I listened to their stories, shared meals, and even participated in farming activities. I saw firsthand the deep connection they had with their lands, their culture, and their community. Their way of life was vastly different, but it was rich and meaningful. This realization hit me hard, shattering my preconceived notions.

I learned that progress and value are subjective. What I had previously viewed as backwardness was, in fact, a different, but equally valid, way of life. The experience made me reconsider the singular notion of progress I held, which was centered around technology and urban development.

Today, I seek admission to the Master of Public Policy program at Harvard Kennedy School not just to learn about policy frameworks and global issues, but to also appreciate the diversity and depth of human experiences. I hope to develop policies that respect and protect these rich cultural and community diversities. This personal transformation inspires my journey toward empathetic and inclusive policymaking.

Essay # 3 [Written by ChatGPT? __ Yes; __No]

The event that fundamentally altered my persona was the devastating takeover of my country by the Taliban. It was as if I were a spectator at a grim theater, observing an unsettling revolution unfold in real time. I bore witness to the swift collapse of the governmental system I had grown accustomed to, and its replacement with a malevolent regime, all in the blink of an eye.

Having been raised during the democratic phase of Afghanistan, I could never have envisioned the return to power of a group mired in antiquated superstitions. This faction, responsible for over two decades of systematic brutality and terror against Afghans, destruction of infrastructure, execution of journalists and social activists, and ruthless massacre of women and children, would now impose their savage code of laws upon us. The nightmare that I had dreaded materialized before my eyes, casting me into a prolonged depression that lasted two months.

Yet, paradoxically, this disheartening episode proved to be enlightening and led to significant personal growth. The shock wave that swept through me instilled a robust sense of duty towards my fellow countrymen, particularly towards vulnerable women and children. It also fostered a resilience in me, a determination that I would endure whatever adversities life might hurl my way and continue to forge ahead.

The psychologist Victor Frankl's insights in his book, "Man's Search For Meaning," are apt in this context. He posited that while external forces might strip away all we have, they cannot take away our freedom to choose our responses to various situations. Having experienced what felt like the nadir of my life, akin to the deep darkness described by Fyodor Dostoevsky in his novel "Notes From Underground," I now understand that progress is not about speed, but rather about perseverance. No matter how arduous my journey becomes, I will not cease to move forward.