

How will ChatGpt/LLM Change the Workplace?

Peter Cappelli

George W. Taylor Professor of Management

Director – Center for Human Resources

What are Large Language Models??

Same basic idea as machine learning: predicting what word comes next – as “auto-complete” does when texting based on what has been texted before.

LLMs see where a word has been used and what words are around it

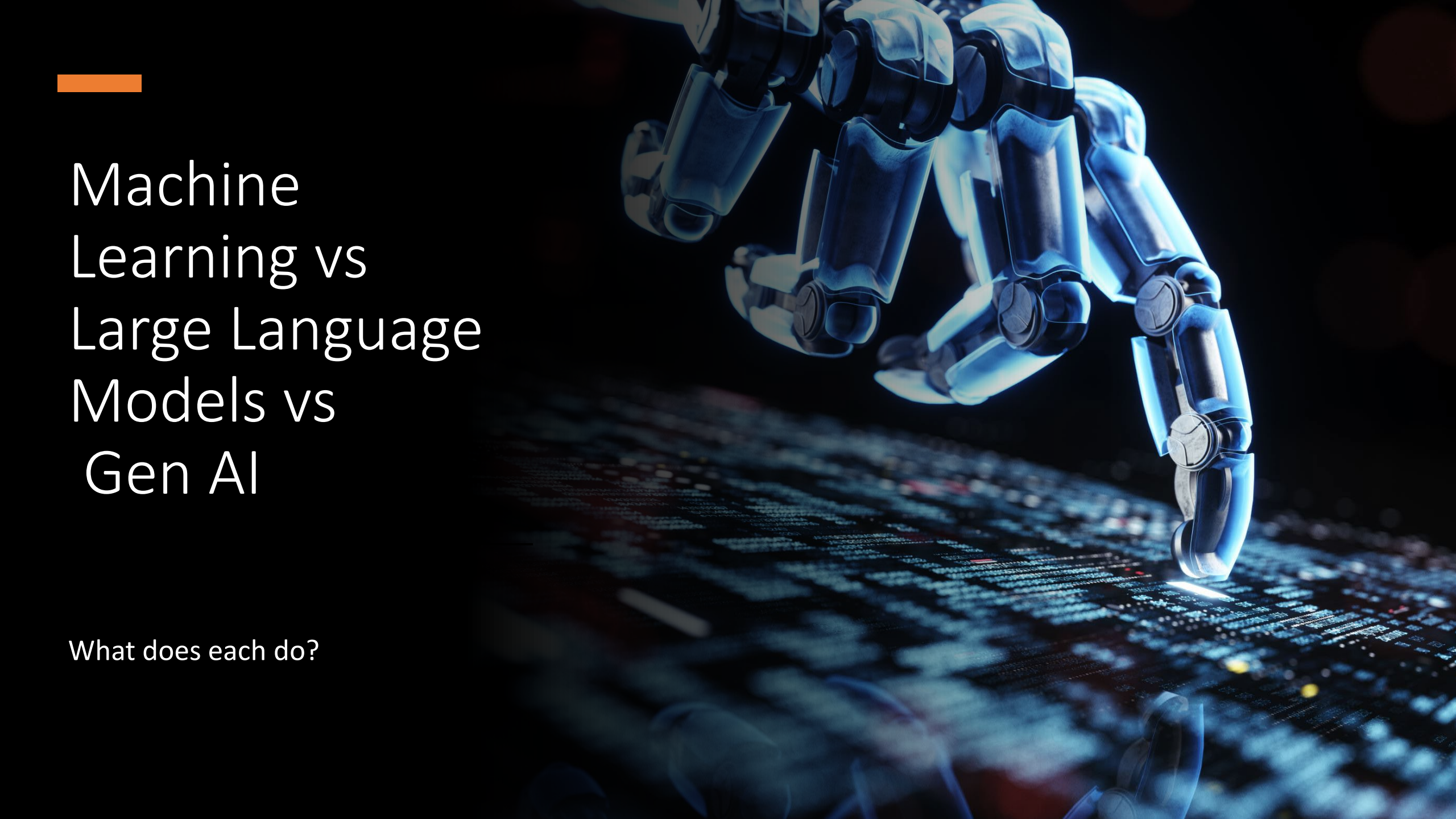
- “Word embedding” = measures how often other words are used around it.
- Do two words mean the same thing? It can tell by measuring the words around them.

“The transformer” – created by Google – does all this for an entire sentence.

- Determines which words are most important to the sentence by looking at all the words at once
- Can tell if the same word is used as a noun or verb. It can figure out what the word “it” refers to in a sentence.
- Extends this beyond individual sentence to paragraphs.
- Predicting the next word in the sequence, then the next, to complete sentences. Possible words get a probability score of likelihood of being next.
- If you want creative responses? Adjust the probability score – “the temperature” – to allow choices that are less probable
- Quirky things: “Hallucinations” – expects words or facts that aren’t there in your context because they were in other texts.
- “Reinforcement Learning by Human Feedback” to fix it.

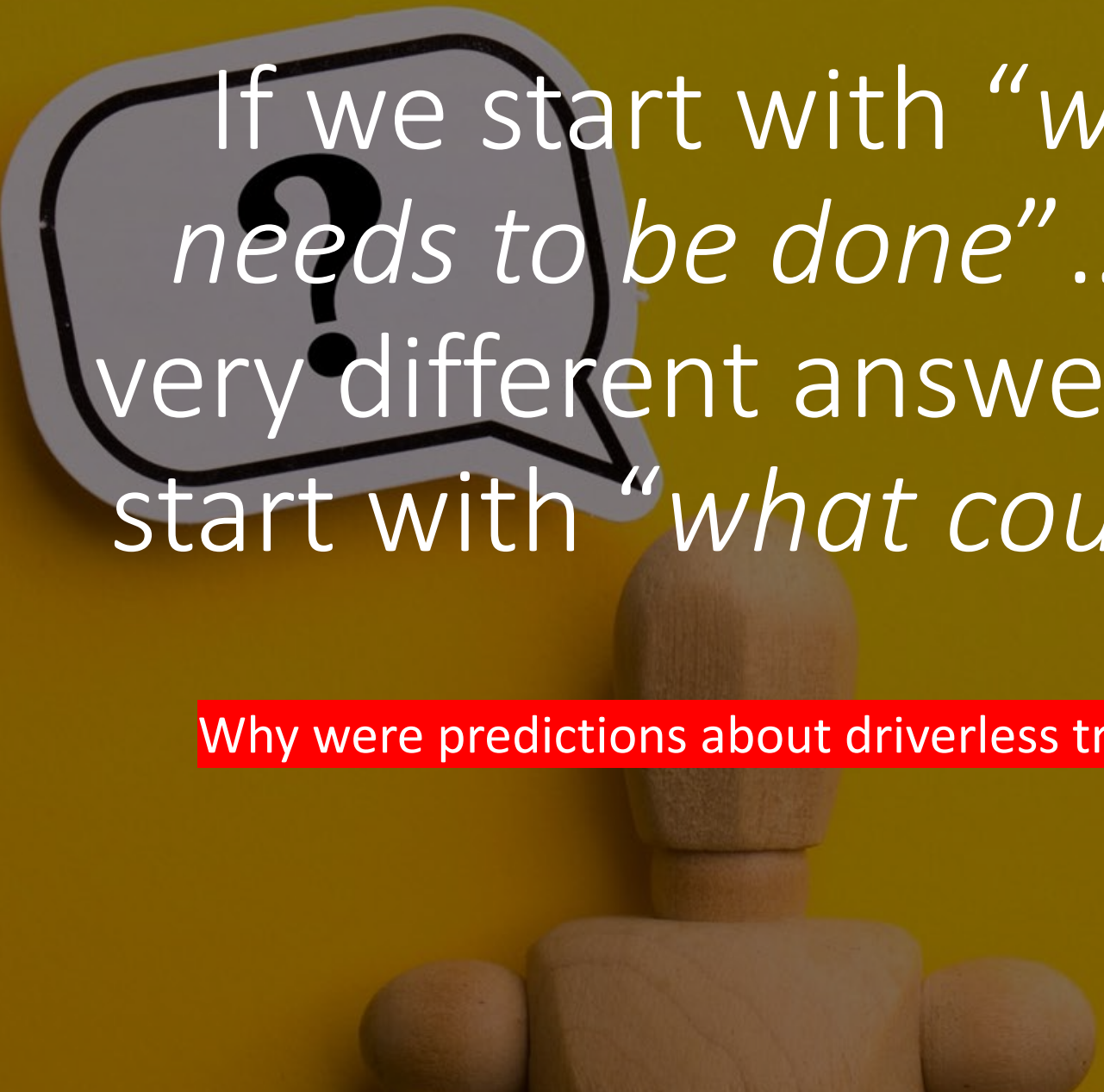
What is remarkable about them

- They can write original-sounding letters, essays, almost anything where the information to do so is on the internet



Machine Learning vs Large Language Models vs Gen AI

What does each do?

A wooden figure is positioned at the bottom center of the frame. To its left, a grey speech bubble with a black outline and a question mark inside is partially visible. The background is a solid olive green color.

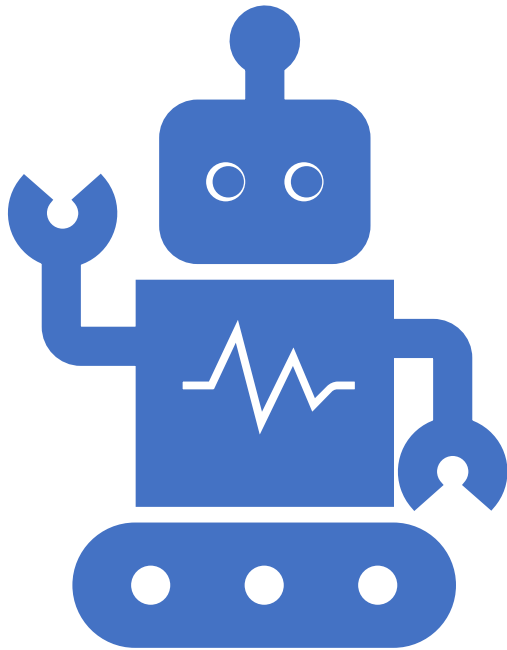
If we start with “*what work needs to be done*” ... we get a very different answer than if we start with “*what could tech do*”

Why were predictions about driverless trucks and cars so off?



Research Evidence:

- Productivity effects are mixed – help in some tasks, actually hurt in others
- When allowed to use Chat Gpt to work on tasks such as writing reports...
- Do regular managers use it or not?
- **Yes.**
- Do they allow Chat Gpt to write a draft and then they edit or rewrite it? Or do they just let Chat Gpt do it and then hand it in?
 - **They just turn it in.**
- Does it save time?
 - **Yes, but so do other solutions such as “form letters”**

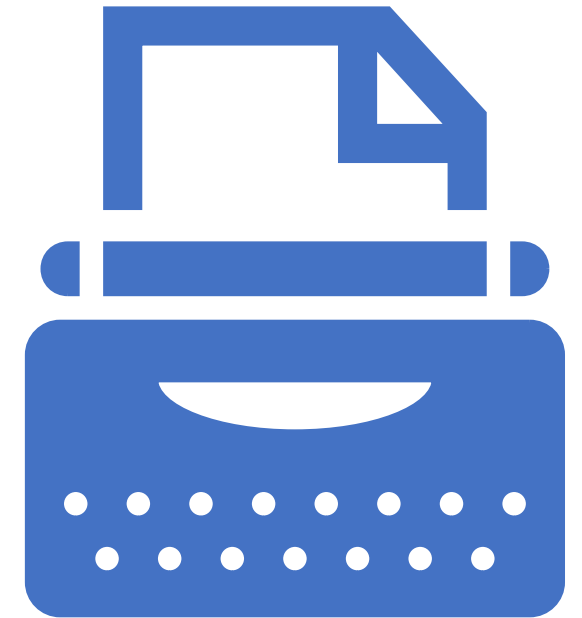


We see three issues with LLMs to think through

- 1st are the ways that Gen AI and LLMs can perform existing tasks worth the change?
 - Q: They could save time writing simple correspondence
 - A: No one writes those now – it's form letters
 - Q: They could do a much better job than regular ChatBots with customer service
 - A: Many organizations don't want "better" customer service.
 - See Air Canada!
 - *It may not be cheap for long, either.*

2nd issue – can we “validate” the output

- LLMs are most useful with novel tasks –
 - Let’s draft a report about something we don’t know!
- Very likely to be better than what we knew before
 - But is it good enough to act on?
 - Google search at least tells you where the evidence comes from
 - It is possible that it is wrong in fundamental ways – can you “validate” the output?
 - Initial evidence is that we don’t check – hard to do, we still need experts





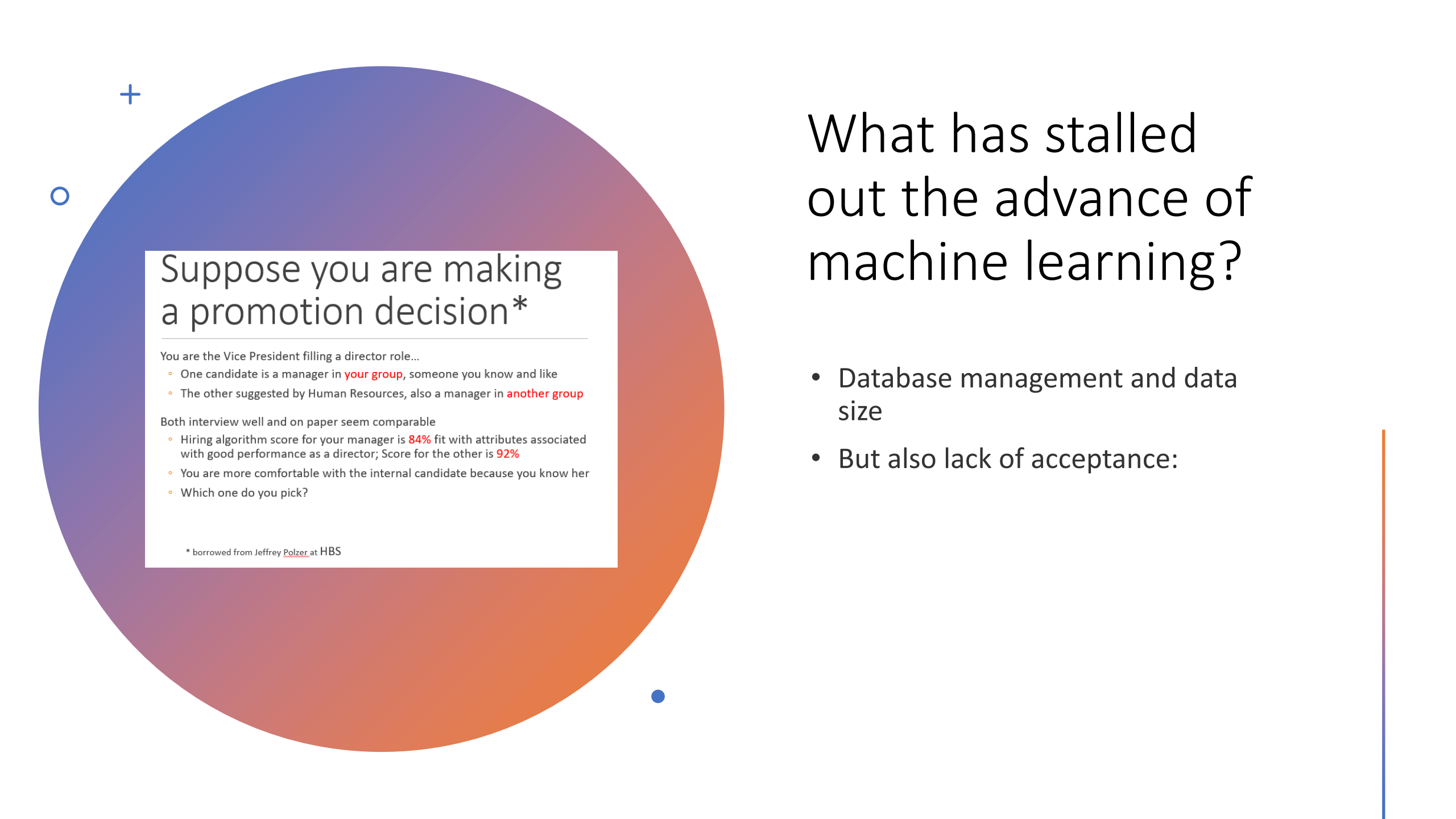
3rd – Can we adjudicate the differences in output

- The same “prompt” on different LLMs will give you different answers
 - The same prompt on the same LLM at a different time can as well
 - Slightly different prompts can give very different answers
- How do we prevent drowning decisionmakers with competing output
 - There is no simply way to explain the differences



What are LLMs really good at?

- Writing term papers
- For entrepreneurs who must act and don't know what to do
 - (A MGMT book might be just as helpful....)
- For analyzing our own data
 - That requires upfront investment
 - "Guard rails"
 - Managing data pollution issues



Suppose you are making a promotion decision*

You are the Vice President filling a director role...

- One candidate is a manager in **your group**, someone you know and like
- The other suggested by Human Resources, also a manager in **another group**

Both interview well and on paper seem comparable

- Hiring algorithm score for your manager is **84%** fit with attributes associated with good performance as a director; Score for the other is **92%**
- You are more comfortable with the internal candidate because you know her
- Which one do you pick?

* borrowed from Jeffrey Polzer at HBS

What has stalled out the advance of machine learning?

- Database management and data size
- But also lack of acceptance:







A woman with long, wavy brown hair is shown from the chest up, wearing a dark jacket. Her hands are raised in front of her, palms facing forward, in a gesture of 'stop' or 'no'. The background is a plain, light color.

Will these LLMs lead to
substantial job losses?

NO

*To cut jobs, it has to take over existing tasks
It is much more likely to address needs that we were not meeting before and
generate new tasks in the process.*

Which of the following describe your organization's use of generative AI? (Select all that apply.)

- a. Individual employees are trying GenAI tools on their own. 
- b. We are experimenting systematically with using GenAI on specific tasks. 
- c. We use GenAI regularly to help with specific tasks. 
- d. GenAI performs some tasks that employees did before. 
- e. Some roles have changed as GenAI has taken over many of their tasks. 
- f. We have eliminated some roles because GenAI performs them. 

Survey from MIT's Work/24 Conference

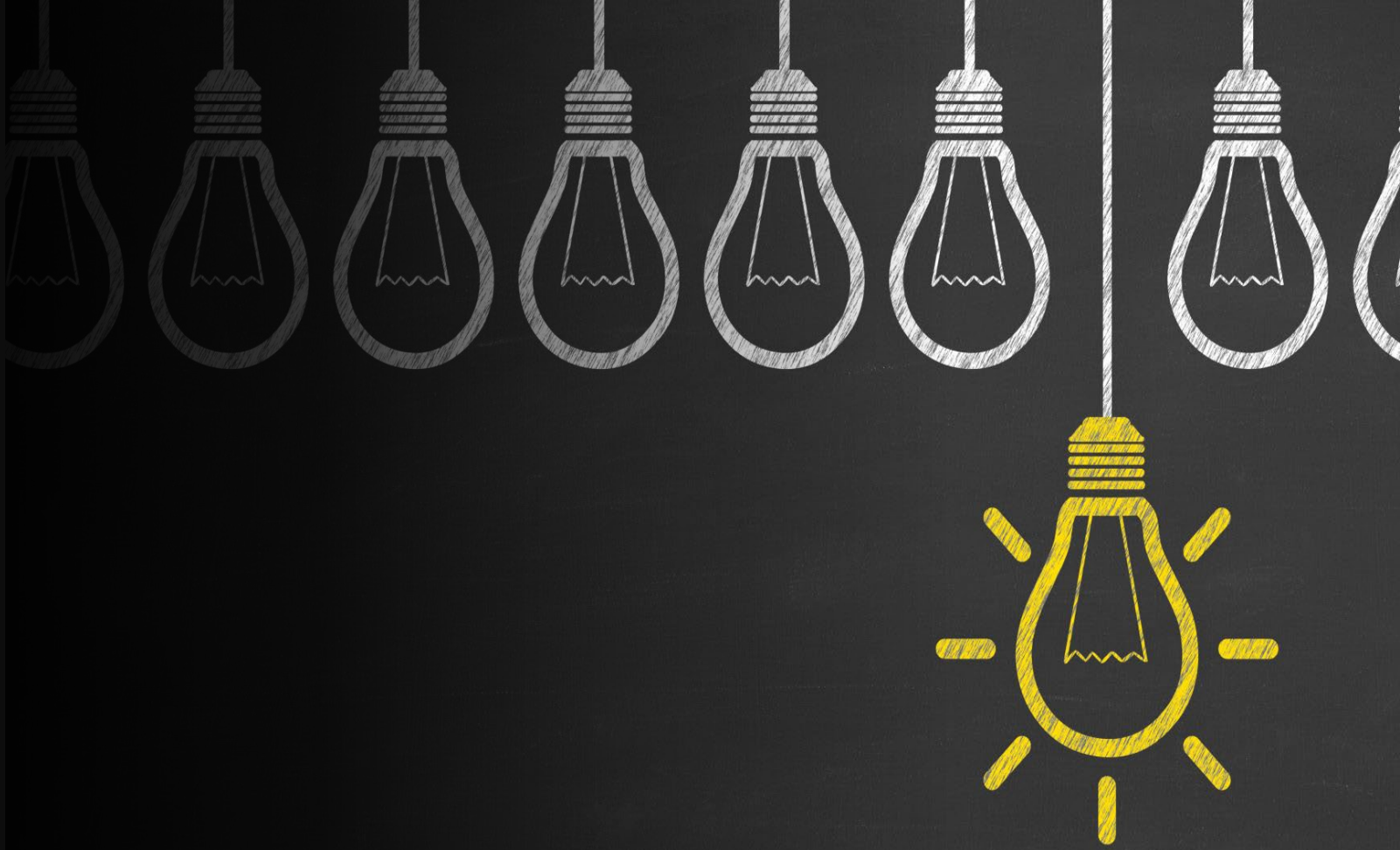
Many reasons why
*“if you build it, they
won’t necessarily come.”*



- Remember – something tasks it could do don’t need doing
- Innovations can only cut # of jobs if it replaces something we already do that takes time.
- Even then, jobs have multiple tasks
- Can’t cut 10% of an employee –
 - Even if it did 1/3rd of a school Principal’s work, could you cut Principals from 1/3rd of the schools??
 - Otherwise, only if you have a typing pool –identical jobs and identical interchangeable employees
- LLMs generate new tasks that someone has to do:
 - Setting guard rails, loading proprietary data, check the output for errors, decide the answer “good enough,” “dueling reports” – adjudication needed because the answers vary

LLMs could be very helpful -

- Like a super search engine –
- Allows us to get better information before acting
- Do new analyses with all the data we don't have now
- Just recognize it is going to be much more interesting to the people that build them than to the people who use them



What is great about them /limitations

- What's Great:
 - Unlike machine learning, they are already built, easy to use, instant results
 - Synthesize big text quickly, like a search engine that writes up the results
- Limitations:
 - To answer questions specific to your organization, they have to be supplemented –no longer free
 - Nonsense on the internet, means nonsense in your answers – need “guardrails,” no longer free
 - Very hard to judge how good the answer is – easiest for simple questions with “good enough” answers
 - Different LLMs produce different answers to the same questions:

How is Gen AI being used right now? (MIT's Work/24 online conference May 2024)

Which of the following describe your organization's use of generative AI? (Select all that apply.)



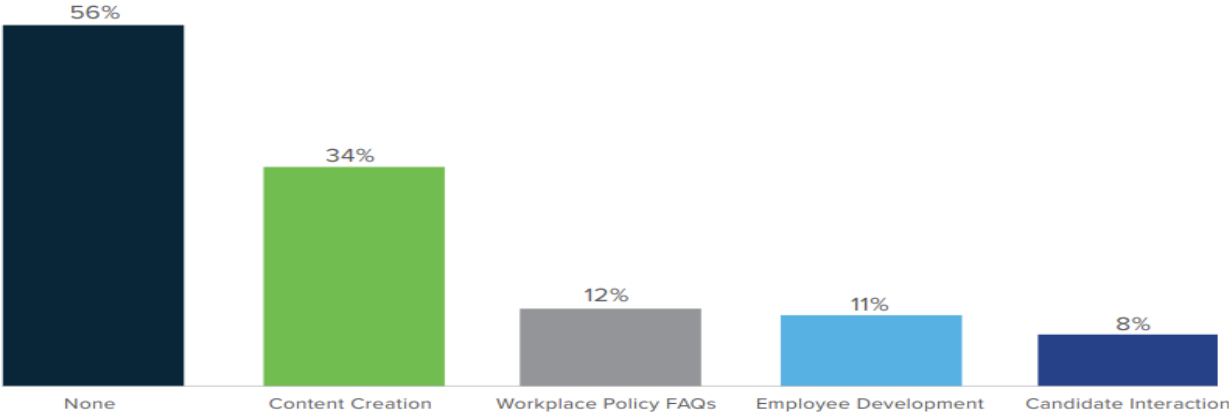
These people are lying :)

AI IN THE WORKPLACE | A LITTLER SURVEY REPORT

GENERATIVE AI USES

Though generative AI technologies have received a lot of attention over the past 12 months, they have not yet penetrated HR departments as deeply as predictive AI. When asked specifically about their organizations' use of generative AI tools, more than half of respondents (56%) say their organizations do not use them in any HR capacity.

In which of the following ways has your organization used generative AI to assist with HR processes? (Select all that apply)





Jump in with a REAL problem

- Given Chat Gpt and other Large Language Models...
 - *What should Universities do now with “term papers?”*

How do we manage ChatGpt in the office?

- For example:
 - You need a recommendation to the top leaders about our business strategy.
 - Specifically, analyze the market opportunities in Canada for our products.
- *Do you let your team use Chat Gpt? Why or Why Not?*
- *Do you let others produce their own competing Chat Gpt report?*
 - *Or do you create a central “office” to produce all these reports?*



Why not just let ChatGpt produce reports?

- What's different about a report for your company vs. a term paper?
 - Company specific information
 - Context – what are the unique concerns in your company, risk-tolerance, culture, etc.
 - We want judgment – Chat Gpt has no judgment
 - *It may well recommend actions that you have already tried because they are “sensible” and have failed for you.*

What to think about - advice

- We need rules: “Acceptable Use Policies”
 - Particularly so when using your own data
 - Acknowledge when it was used
 - Certify that output has been checked
- Training to use LLMs is important
 - Really important to understand the limitations: We won't get this just by practicing
- Do we need a central office to conduct LLM analyses and reports?
 - Manage the data, manage validity problems, adjudicate differences





In summary

- It is a fabulous tool –
 - It is going to be more work than we think to use in all but simple tasks
 - It generates useful information
 - But not necessarily answers we can act on –
 - We have to manage the flow of that information
 - We have to process it to make judgments