

LEGAL USE OF LARGE LANGUAGE MODELS

MODERN BAR ASSOCIATION:

NEW LEGALTECH CHALLENGES - LEGALTECH IN EUROPE -

POLAND, WARSAW, 15-16 JUNE



Peter Homoki, homoki.net

LLM as a technique: what makes this exciting now?

Opportunities and limits for lawyers

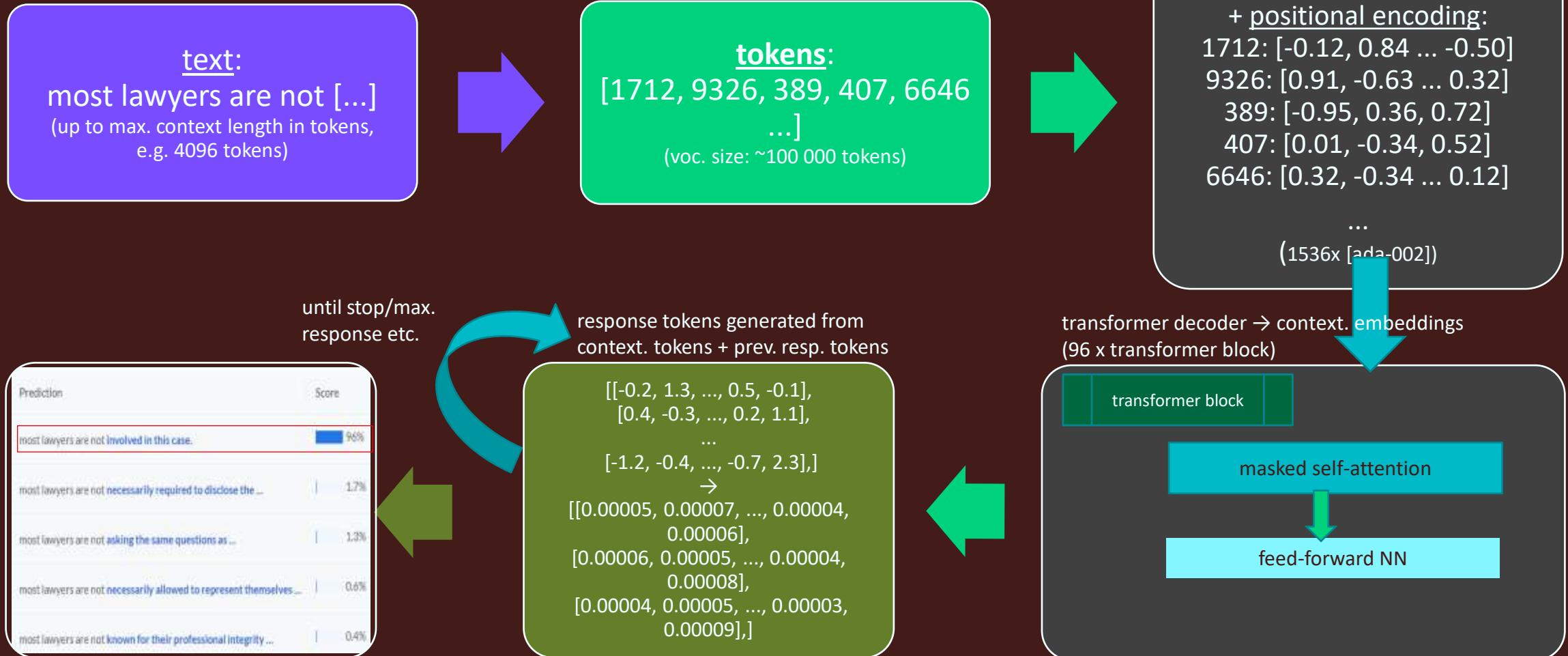
Risks and human in the loop?

Competitive edge among lawyers using LLMs

CCBE/ELF Guide on the use of AI for lawyers (2022): <https://ai4lawyers.eu/>
See some the examples in more detail at my blog: <https://homoki.net/en/blog/>



LLM: a language probability model



LLMs became exciting because ...

emergent abilities: with *appropriate architecture* and a large enough training data set...

new, surprising abilities appeared with

the increase of parameters (neural network connections)

such as ...

- prompt-based tasks, in-context learning (versatile/practical uses in diverse zero-shot/few-shot tasks, w/o fine-tuning)
- better logic reasoning, better understanding of human input, e.g. can translate without being taught to

instruction tuning (fine-tuning) as a generic way to improve completion of new tasks, even cross-languages

public awareness of the capabilities of AI tools since November 2022

Opportunities for lawyers with LLMs

Simplified, generic uses of machine learning made possible:

- number of usable experts ↑ (e.g. consultants, and not ML engineers needed)
- no costly data preparation or fine-tuning needed (implementation costs ↓)
- better understanding of own ML use cases

New, accessible AI-based applications in general

- replacement of less reliable, complex software (e.g. grammar generation for multiple languages)
- small number of tools to integrate

Use of conversational UIs (chat, open book question-answering) in legal practice

As of now, these tools can be an equalizer between large and small firms

The following is a list of companies and the categories they fall into:

Apple, Facebook, Fedex

Apple
Category:

Engine	text-davinci-003
Max tokens	64
Temperature	0
Top p	1.0
Frequency penalty	0.0
Presence penalty	0.0

Sample response

Technology

Facebook
Category: Social Media

Fedex
Category: Delivery

Prompt

Extract keywords from this text:

Black-on-black ware is a 20th- and 21st-century pottery tradition developed by the Puebloan Native American ceramic artists in Northern New Mexico. Traditional reduction-fired blackware has been made for centuries by pueblo artists. Black-on-black ware of the past century is produced with a smooth surface, with the designs applied through selective burnishing or the application of refractory slip. Another style involves carving or incising designs and selectively polishing the raised areas. For generations several families from Kha'po Owingeh and P'ohwhóge Owingeh pueblos have been making black-on-black ware with the techniques passed down from matriarch potters. Artists from other pueblos have also produced black-on-black ware. Several contemporary artists have created works honoring the pottery of their ancestors.

Settings

Engine	text-davinci-003
Max tokens	60
Temperature	0.5
Top p	1.0
Frequency penalty	0.8
Presence penalty	0.0

Sample response

Keywords: Black-on-black ware, 20th century, 21st century, Puebloan Native American ceramic artists, Northern New Mexico, reduction-fired blackware, pueblo artists, selective burnishing, refractory slip, carving/incising designs, polishing

SYSTEM

You are a silent AI model changing the grammar and the name (category) of subjects according to the following general instructions:

{acc.} means change the preceding term into accusativus case,

{dat.} means change it to dativus, {gen.} means change it to

genitivus.,

{sing.} means leave the preceding term in singular, {plur.} means

change the preceding term to plural.

Delete all terms in brackets after change.

Further instructions are set in "Change required:" part of the user instructions between brackets "{}", and within the brackets, multiple instructions are separated by quotation marks and comma ("", "").

The text to be converted follows the last closing bracket "}".

USER

Change required: {"replace [Customer] with Bérbeadó", "replace [Contractor] with Bérelő", "replace [Subject] with Bérlemény"}

A [Contractor]{plur.} a Bérleti Idő teljes tartamára a [Customer]{sing.} részére bérbeadja{plur.} az Épületben{plur.} található alábbi alapterületű [Subject]{acc.}{plur.} és a hozzá{gen.} tartozó parkoló{acc.}{sing.}, valamint kifizeti a [Customer]{dat.}{sing.} a Díj{acc.}{plur.}

ASSISTANT

A Bérlők a Bérleti Idő teljes tartamára a Bérbeadó részére bérbeadják az Épületekben található alábbi alapterületű Bérleményeket és a hozzájuk tartozó parkolót, valamint kifizetik a Bérbeadónak a Díjakat.

⊕ Add message



integration into document assembly/automation tools:

- a) **clause** level (existing and practical as of now)
- b) contract **structure** level (research)
- c) contract **workflow** process level: review, negotiation, playbooks, checklists (research)

question	answer	Human answer considered as correct	Human evaluation
Ha már eljárt az ügyben kirendelt igazságügyi szakértő, mikor rendelhet ki a bíróság új igazságügyi szakértőt?	A bíróság akkor rendelhet ki új igazságügyi szakértőt, ha a kirendelt szakértő szakvéleménye aggályos és az aggályosság a szakértő által adott felvilágosítás ellenére sem volt kiküszöbölhető, ahogy az a 315. § (1) bekezdésében olvasható.	Ha a kirendelt szakértő szakvéleménye aggályos, az aggályosság a szakértő által adott felvilágosítás ellenére sem volt kiküszöbölhető és a fél ezt indítványozza.	1
Mikor lehet válasziratot beterjeszteni?	A válasziratot az anyagi pervezetése szerinti felhívásra vonatkozó nyilatkozatot követően lehet beterjeszteni. A fél válasziratban külön felhívás nélkül is tehet egyéb perfelvételi nyilatkozatot, és a 199. § (5) bekezdése szerinti felhívást intézhet az ellenfélhez. Az előkészítő iratban - a (2) bekezdésben foglalt kivétellel - azt a perfelvételi nyilatkozatot kell feltüntetni, amelynek ilyen formában történő előterjesztésére a bíróság a felet felhívta.	Ha a bíróság erre a felet felhívta, illetve a kereset-változtatás, viszontkereset-változtatás és ellenkérelem-változtatással összefüggésbenálló válasziratot	1
Hány éves kortól korlátozottan cselekvőképes egy személy?	A törvény szerint a korlátozottan cselekvőképes személyek kiskorúak és nagykorúak is lehetnek. A kiskorúak cselekvőképtelensége a tizennegyedik életévük betöltéséig tart, míg a nagykorúak esetében a cselekvőképesség részleges vagy teljes korlátozása a bíróság határozatával történhet.	14 éves kortól	1
Mi a dolog?	A dolog a tulajdonjog tárgya, és egy birtokba vehető testi tárgy lehet . A pénzre, értékpapírokra és a dolog módjára hasznosítható természeti erőkre is vonatkoznak a dologra vonatkozó szabályok. Az állatokra is alkalmazni kell a dologra vonatkozó szabályokat, figyelembe véve a természetüknek megfelelő eltéréseket.	Birtokba vehető testi tárgy	1
Mi az ingatlan?	Az ingatlan egy olyan terület, amelyen épület, építmény vagy más módon hasznosítható ingó dolog található. Az ingatlan tulajdonjoga kiterjed az épületre, az építményre és a hozzá tartozó földre is. Az ingatlanok lehetnek lakóházak, irodaházak, üzletek, raktárak, mezőgazdasági területek	a dolgok természeti tulajdonságuk szerinti egyik fő csoportja az ingó mellett, elsősorban a földterület és a föld méhe, és mindaz, amit a földön az ember vagy a természet létrehozott (a föld természetes vagy mesterséges alkotórésze).	1

<https://homoki.net/en/2023/05/29/On-the-importance-of-lawyer-focused-QA-benchmarks.html>

COLIEE

Is an airline liable for its pilot's negligence?

A carrier would not be liable for an error of judgment of the pilot, not constituting positive negligence on his part in exercising such judgment; but liability is incurred if the pilot, by his negligent and careless conduct, has created a situation requiring the formation of a judgment and then errs in the exercise thereof.

An airline corporation is not an insurer of the safety of its passengers. The liability of an airline corporation must be based on negligence.

Airline pilot who was accused of raping flight attendant has no tort claim against airline based upon its alleged negligent investigation of accusation, even if airline's policy of investigating sexual harassment complaints creates duty to use due care in conducting investigation....

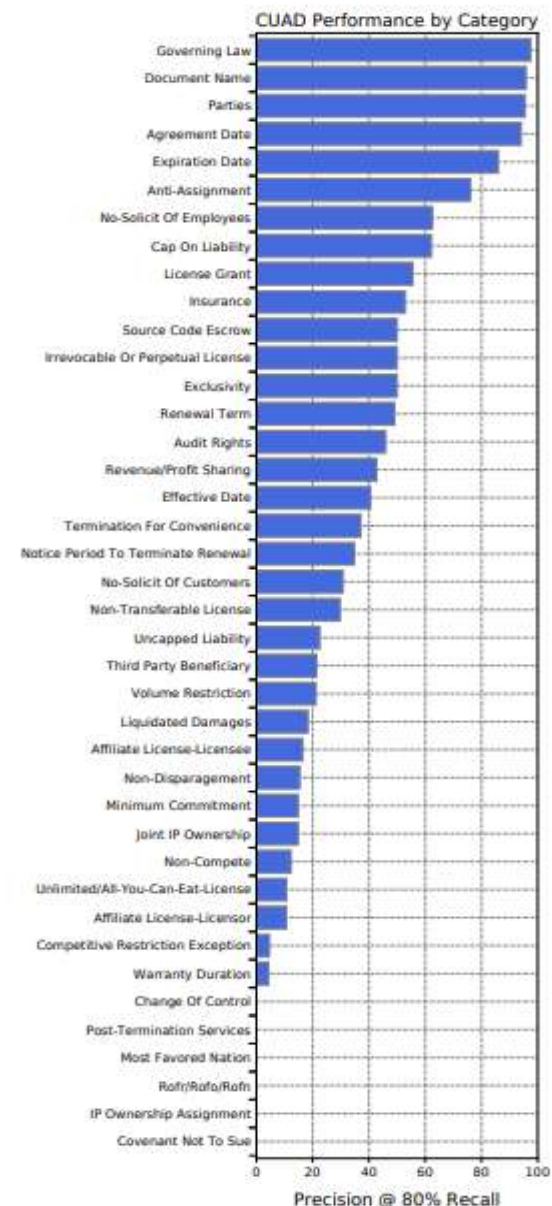
SME label Method

3 AF on BM25_MLT

1 L_GloVe, SL_BERT

0 BM25_MLT

Method	DCG@3 ^a	95% C.I. ^b	N silly ^c	Answered ^d	MRR@3
BM25_MLT	4.052	-	7	100	0.411
SL_BERT	3.386	1.26	2	100	0.326
L_GloVe	2.855	1.25	7	100	0.285
AF BM25	5.464	1.43	7	100	0.493
AF SL_BERT	4.862	1.43	0	100	0.416
AF L_GloVe	4.281	1.40	7	100	0.397
AF (BM25, SL_BERT)	5.605	1.47	5	100	0.483
AF (BM25, L_GloVe)	5.502	1.47	8	100	0.481
AF (BM25, SL_BERT, L_GloVe)	5.533	1.45	6	100	0.492
AF 0.2 (BM25, SL_BERT)	6.269	1.52	2	89	0.543



Limits of usability of LLMs

Language limitations outside English: GPT-3.5 and GPT-4 is usable

- open source small models (LLaMA-based, GPT-J etc.): gen. English only, some national specific
- Cohere: English only etc.
- Google Bard, PaLM 2: English only
(although excellent translation capabilities for many languages are present)
- national huge models? ([„PULI” GPT-3SX \(6.7B\) Hungarian](#) → 100B+ in preparation)

Other technical limitations:

- accessible via API only (risks of confidentiality)
- slow APIs, context size limits (e.g. 3k in CEE instead of 8k/32k, MS Azure version probably better)
- no fine-tuning beyond GPT-3

Caveat emptor

- **errors** even in the best models: hallucinations (in generative models) or other errors
- **use of inappropriate models**: extra dangers in using non-stochastic models, incorrect parameter settings (GPT-4 is not necessarily the best solution for every problem)
- **human in the loop** as a liability shield:
 - ~ „full self-driving“ with mandatory hands on

The currently enabled Autopilot, Enhanced Autopilot and Full Self-Driving features require active driver supervision and do not make the vehicle autonomous. Full autonomy will be dependent on achieving reliability far in excess of human drivers as demonstrated by billions of miles of experience, as well as regulatory approval, which may take longer in some jurisdictions. As Tesla's Autopilot, Enhanced Autopilot and Full Self-Driving capabilities evolve, your car will be continuously upgraded through over-the-air software updates.

Human (lawyer) in the loop

“human in the loop” is *not a liability shield for providers*, but a *principle for use cases*:

- now we are still in the research phase
- focus of LLM use should be on B2B legaltech solutions, not B2C (direct-to-consumer) solutions (review by lawyer)

difficulties:

- chaining of tools
- **human bottleneck**: acceptance that there is no way around it for many years → AI tools **to improve quality of service, not quantity**
- some tools will not be used, even if they are technically available

until we are convinced of the reliability of the tool in practice (like self-driving cars on the Autobahn)



Chip- and HW manufacturers
 Large commercial LLM providers.
 (Google, MS/OpenAI etc.)



Research institutes/universities



Cloud-based IaaS/PaaS providers



Open source consultants, consultants of open language models (OLM), providers of non-proprietary AI solutions



Owners of private data



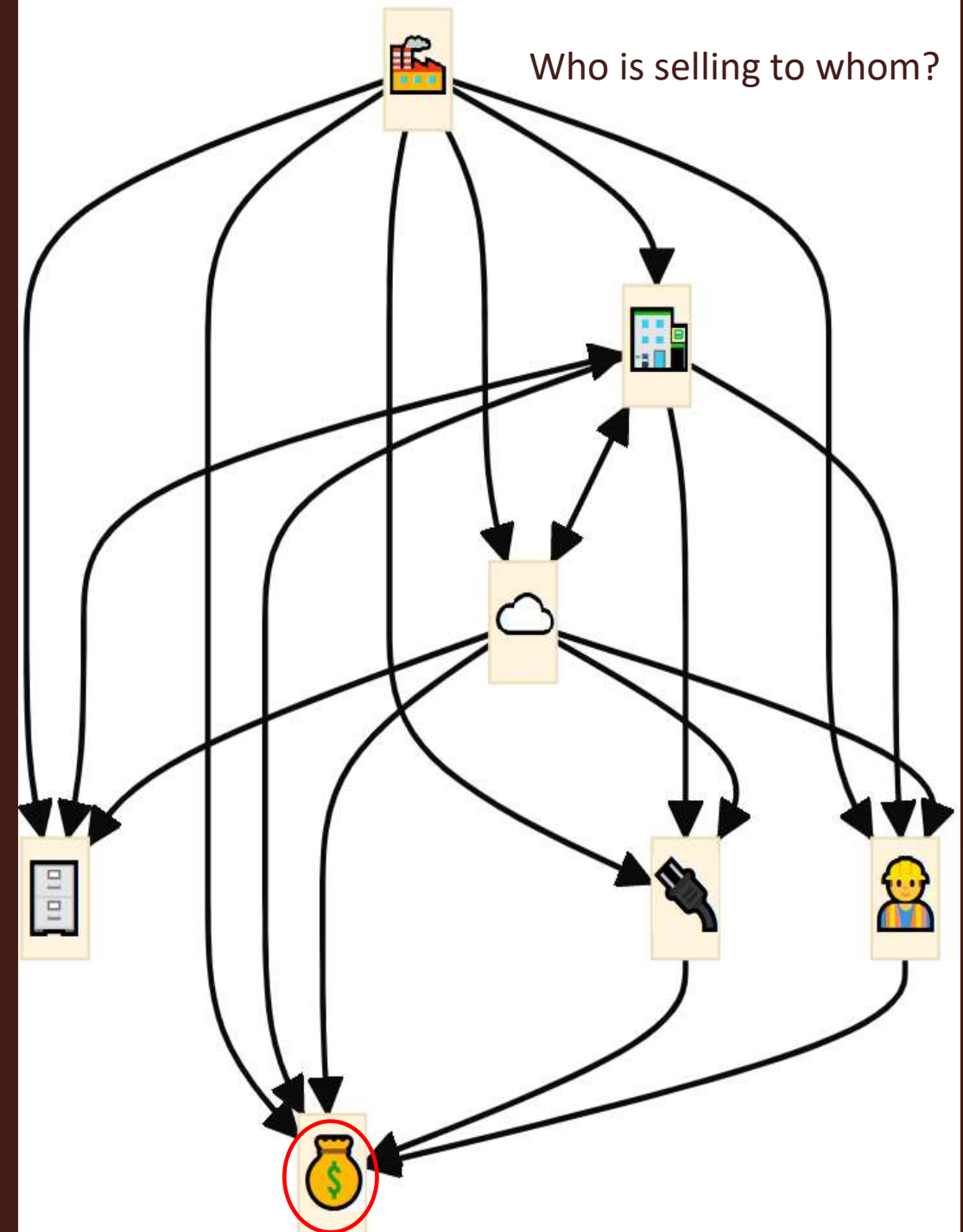
Customers



AI “downstream” providers:
 plugin providers, sellers of fine-tuned models,
 SDK/API



Governments



Competitive edge among lawyers using LLMs in an AI ecosystem

Custom capabilities	Knowledge	Processes	Client reach, trust, distribution
Custom application	Using custom models	Using diff. APIs (chains etc.)	Relying on other custom applications
Custom model	Model trained on custom data: task-specific NN train, pre-train, fine-tune, LoRA ...	Model providing better performance	Model w/ stronger control over data
Custom data	Client data	Process data	Knowledge base data



Thank you!

H-1137 Budapest

_Katona József utca 39._4/22.

tel. +36(1) 700 4020

email office@homoki.net

homoki.net



_Homoki
Law
Firm

Suggested sources for further reading:

PaLM 2 report: <https://ai.google/static/documents/palm2techreport.pdf> (May 2023)

GPT-4 “Sparks of AGI”: <https://arxiv.org/abs/2303.12712>

GPT-4 technical report: <https://arxiv.org/abs/2303.08774>

On measuring emergent abilities (BIG-bench): <https://arxiv.org/abs/2206.04615>

Open source models: https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard

More generic AI landscapes: <https://www.antler.co/blog/generative-ai>,
<https://www.sequoiacap.com/article/ai-50-2023/>

homoki.net/en/blog