Pearson

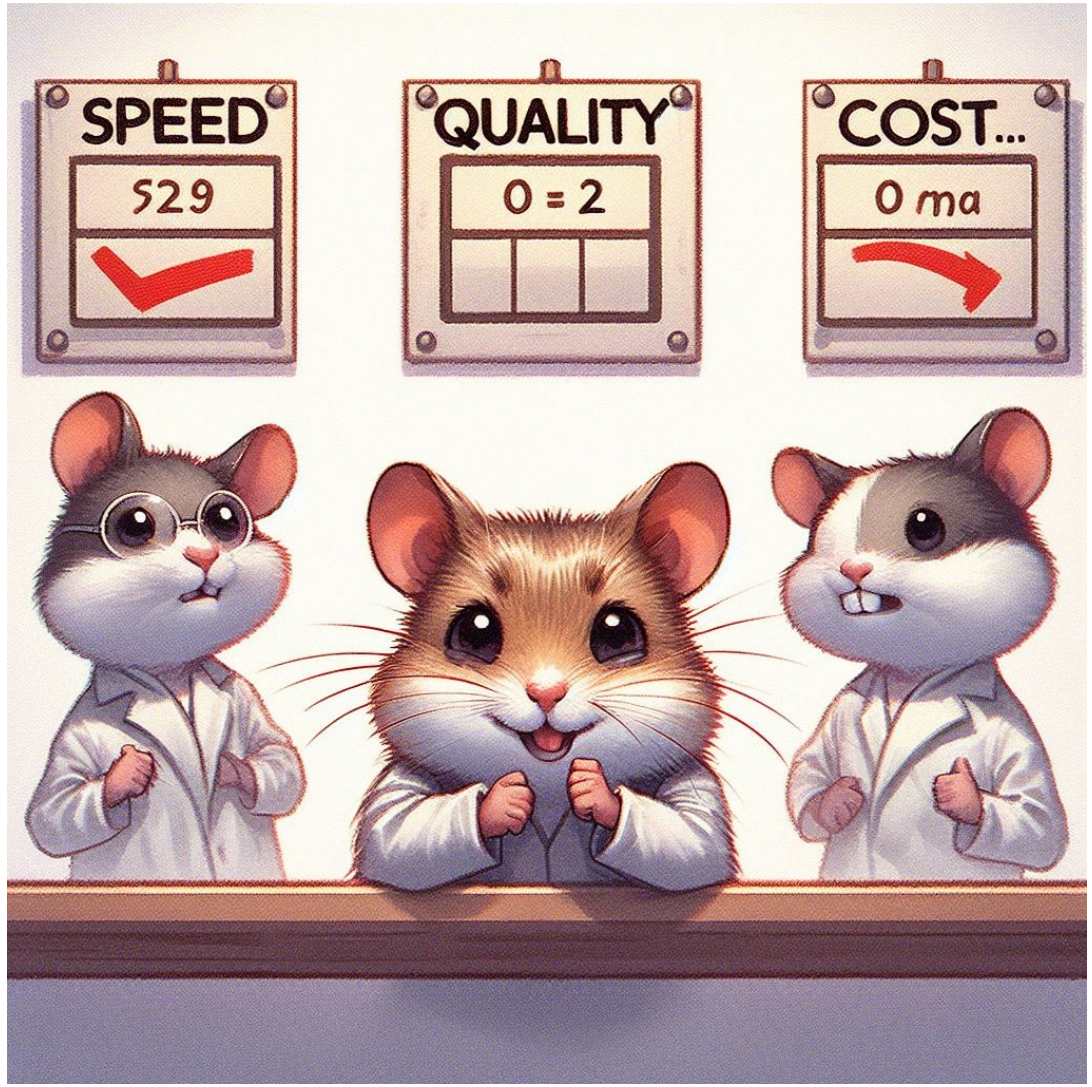
EdTech (R)Evolution: LLM Unleashed - Balancing Speed, Quality, Costs and Ethics

6.04.2024
Krzysztof Sopyla (krzysztof.soopyla@pearson.com)

Pearson

Going deeper into the rabbit hole

Pick the two?



ELLTech

Agenda

1. **Our use cases**
   The classes of problems we are working on.
2. **LLM definition**
   How to define LLMs, what are the key characteristics
3. **Align with human values**
   The missing piece of the puzzle
4. **Key areas to optimization**
   What to consider when implementing LLM efficiently.
5. **Model deployment**
   What things we should take into consideration.
6. **Thank You**

AI Education
Our North Star

Content Provider

It provides access to educational materials, extracts ready-made materials or generates them tailored to student interests, region and level.

AI Tutor

It is leading the lesson, managing the conversation flow, and sticking to the lesson plan.
Strives to achieve the student's educational goals

Response Assessment

The Student's Response Assessment Model gives feedback on what he did well and what he did wrong. Assesses pronunciation, grammar, vocabulary, etc.

Learner Model

A student profile. Processes all information about students to compute their proficiency in particular domains. Recommends what to learn next.

Use case

Conversational AI

Education copilots (passive), Speaking practice assistants (free and role play), AI tutoring

Use case

Content generation

English learning materials: readings, the script for listening, grammar and vocabulary activities

Use case

Safeguarding and LLM Evaluation

Fast topic recognition, Content Policy violation, Pearson tailored Evaluation

LLM – key characteristics

How can we define LLM?

L – large ?

- # of parameters (>1B)
- most often identified as a decoder from the Transformer architecture (GPT family), encoder-decoder or RNN (RWKV model)
- Scaling laws for neural language model 2020 OpenAI

LM – language model

- a model that has encoded knowledge about the structure of the language, its semantics,
- can generate coherent and understandable text
- "A neural probabilistic language model" Y. Bengio at all, 2003

Has knowledge

- has knowledge about culture, history, exact sciences, etc., at least partial
- knows how to use this knowledge and connect the facts
- "Survey on Factuality in Large Language Models: Knowledge, Retrieval and Domain-Specificity", 2023

Knows how to reason

- has the ability to aggregate and summarise information
- has the ability to reason logically based on the information provided
- Emergent Abilities of Large Language Models, 2022

LLM – key characteristics

How can we define LLM?

- **L – large**
- **Language Model**
- **Has knowledge**
- **Knows how to reason**

Align with human values

- [OpenAI Superaligment program]
- Eastern vs Western Culture
- Underrepresented cultures and values

AI ethics
Wyszukiwane hasło

AI threat
Wyszukiwane hasło

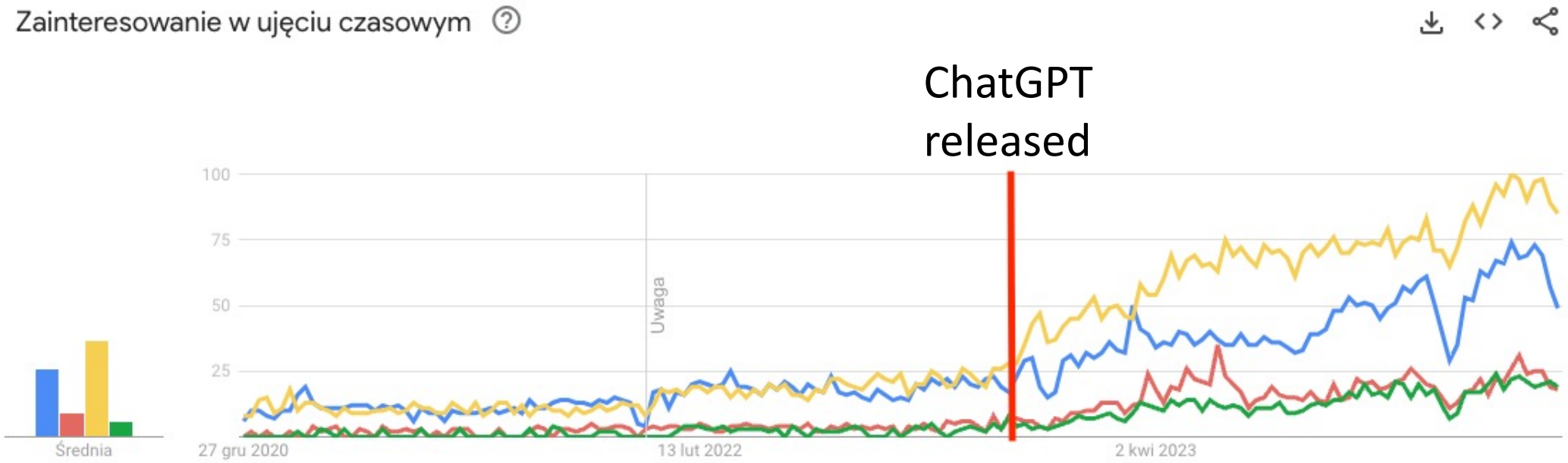
AI speed
Wyszukiwane hasło

AI costs
Wyszukiwane hasło

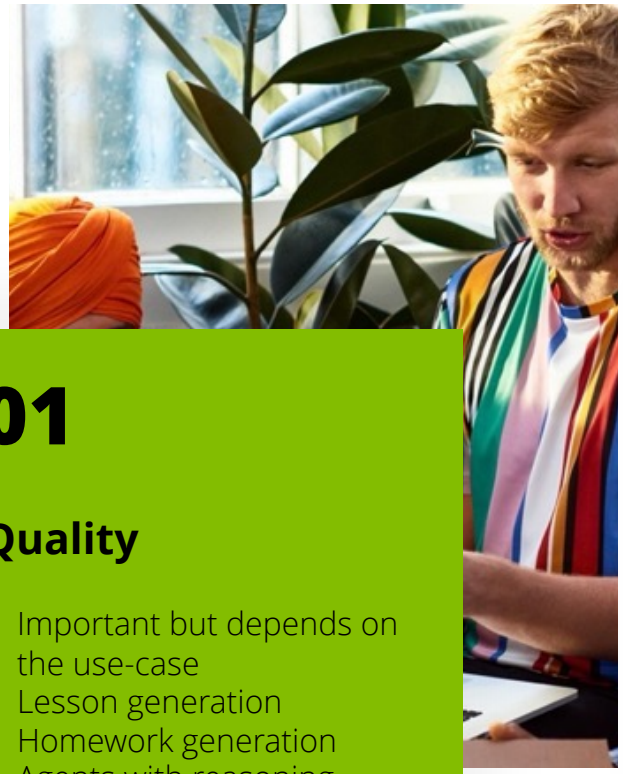
+

Cały świat | Od 1.01.2021 do 2.04.2024 | Wszystko | Wyszukiwarka Google

Zainteresowanie w ujęciu czasowym



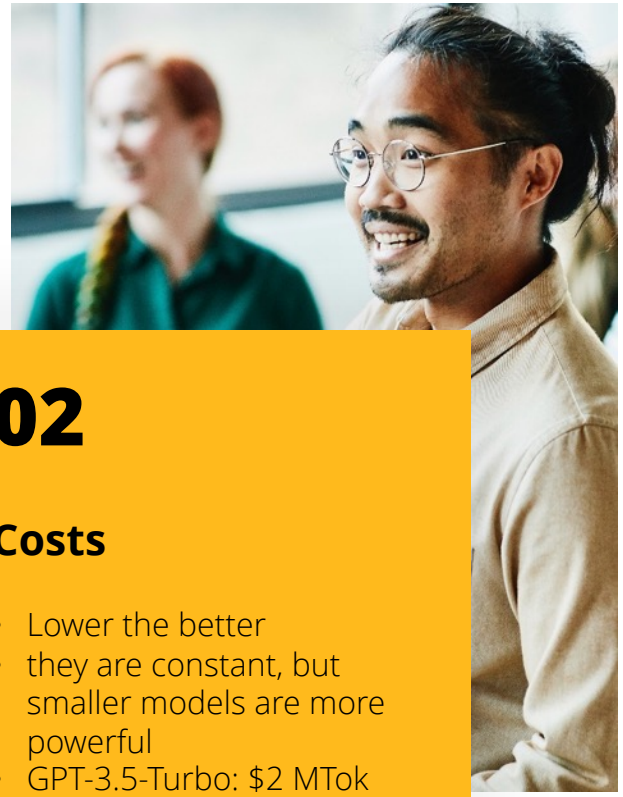
Key criteria?



01

Quality

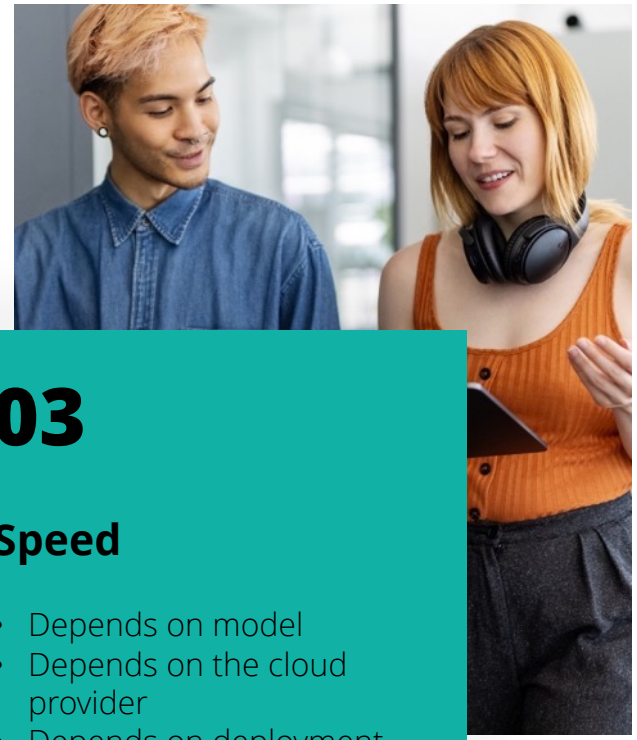
- Important but depends on the use-case
- Lesson generation
- Homework generation
- Agents with reasoning



02

Costs

- Lower the better
- they are constant, but smaller models are more powerful
- GPT-3.5-Turbo: $2 MTok
- GPT-4-Turbo: $40 MTok
- GPT-4: $90 MTok
- Claude Opus: $90MTok



03

Speed

- Depends on model
- Depends on the cloud provider
- Depends on deployment server (TGI vs vLLM)
- Depends on use-case: chat vs content gen

Deployment aspects

When planning to implement an LLM-based product, the following aspects should be considered

What to consider?	Content generation	AI tutor
Model size	Medium, Large	Small, Medium
Deployment type	API LLM providers (OpenAI, Azure, AWS Bedrock, Claude etc)	Custom TGI or similar,
Costs	small to medium, store generated content and reuse it	in scale they can be large, so custom deployments are prefered
Quality	very important, low hallucination, stick to the facts	medium importance in conversation, especially when it is paired with guardrails
Speed	slow to medium, we can hide this with UI	Very fast, mainly when you communicate by voice
Regionalisation, localisation	important	moderately important
Safety	not so important, often there is human in the loop	very important, there is no time to human review
Leagal aspect	important, pick the right partner, have information about the data license that was trained on	important, check the license of the Open-source model
Model replacement, refreshing	moderately important, if you choose one of the main provider you should be ok	very important, prepare for model changes, fine-tuned version on your own data

Pearson

Thanks for your attention

Let’s stay in touch

Krzysztof Sopyła

- krzysztof.sopyla@pearson.com
- https://www.linkedin.com/in/krzysztofsopyla/

Pearson LinkedIn – job offers

- https://www.linkedin.com/company/pearson/

ELLTech