Using AI to Detect the Unusual

*Introduction to AI and anomaly detection — **FARI Happy Hour***

VUB ARTIFICIAL INTELLIGENCE RESEARCH GROUP

1. **Using AI** to detect the unusual
2. Using AI to **detect the unusual**
3. **Using AI to detect the unusual**

1. **Using AI** to detect the unusual
2. Using AI to **detect the unusual**
3. **Using AI to detect the unusual**

Artificial Intelligence

- **Field of research:**
  - Branch of **mathematics** and **computer science**
  - Interdisciplinary by nature

- **Since 1950s** – increase in computing power allows us to achieve remarkable things!

Using AI

- **Field of research:**
  - Branch of **mathematics** and **computer science**
  - Interdisciplinary by nature

- **Since 1950s** – increase in computing power allows us to achieve remarkable things!
  - **Example:** Search

Deep Blue vs. Kasparov

Deep Blue
IBM chess computer

Garry Kasparov
World Chess Champion

Deep Blue versus Garry Kasparov was a pair of six-game chess matches between then-world chess champion Garry Kasparov and an IBM supercomputer called Deep Blue. Kasparov won the first match, held in Philadelphia in 1996, by 4–2. Deep Blue won a 1997 rematch held in New York City by 3½–2½. The second match was the first defeat of a reigning world chess champion by a computer under tournament conditions, and was the subject of a documentary film, *Game Over: Kasparov and the Machine*.

Sources:

1. https://en.wikipedia.org/wiki/Deep_Blue_versus_Garry_Kasparov

Using AI

- **Field of research:**
  - Branch of **mathematics** and **computer science**
  - Interdisciplinary by nature

- **Since 1950s** – increase in computing power allows us to achieve remarkable things!
  - **Example:** Search

AlphaGo versus Lee Sedol, also known as the **DeepMind Challenge Match**, was a five-game Go match between top Go player Lee Sedol and AlphaGo, a computer Go program developed by DeepMind, played in Seoul, South Korea between 9 and 15 March 2016.

Go is a complex board game that requires intuition, creative and strategic thinking.[8][9] It has long been considered a difficult challenge in the field of artificial intelligence (AI). It is considerably more difficult[10] to solve than chess. Many in artificial intelligence consider Go to require more elements that mimic human thought than chess.[11]

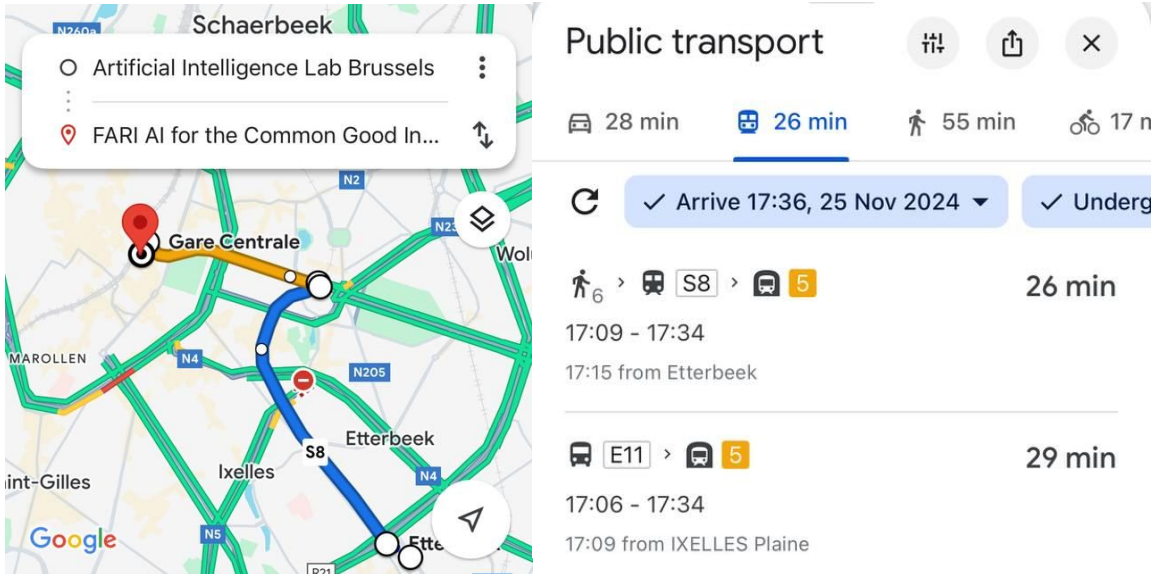
Sources:

1. https://en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol

Using AI

- **Field of research:**
  - Branch of **mathematics** and **computer science**
  - Interdisciplinary by nature

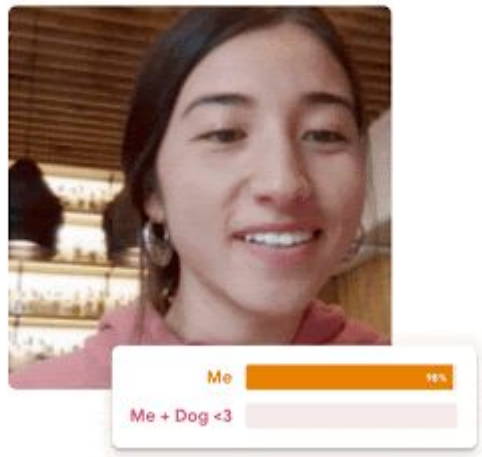
- **Since 1950s** – increase in computing power allows us to achieve remarkable things!
  - **Example:** Search
    - *You might not see this as "AI", but results from AI research!*

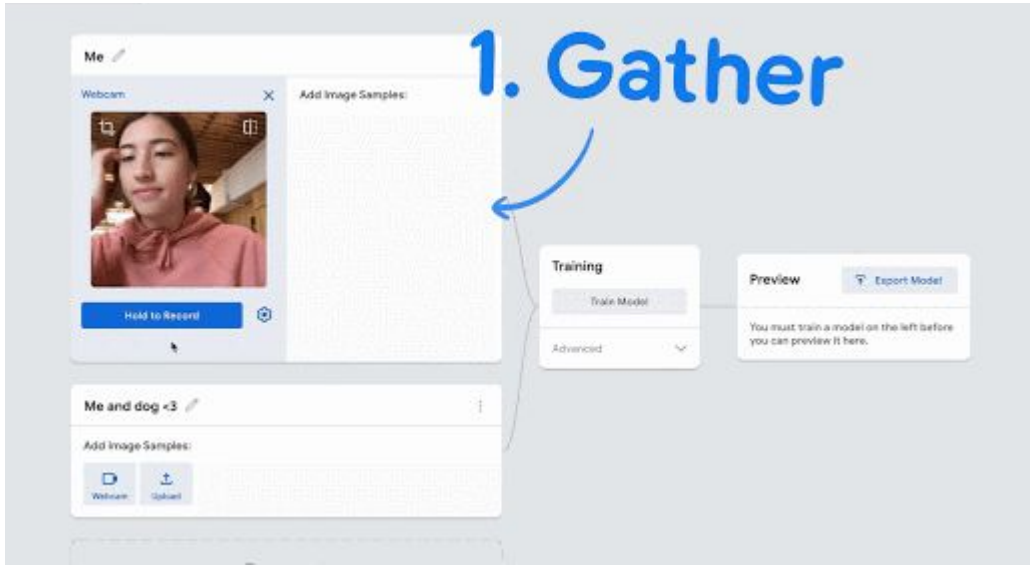


Using AI

- **Field of research:**
  - Branch of **mathematics** and **computer science**
  - Interdisciplinary by nature

- **Since 1950s** – increase in computing power allows us to achieve remarkable things!
  - **Example:** Classification
    - *Labelled images of cats vs. dogs*
    - *Assigning a label to new images*





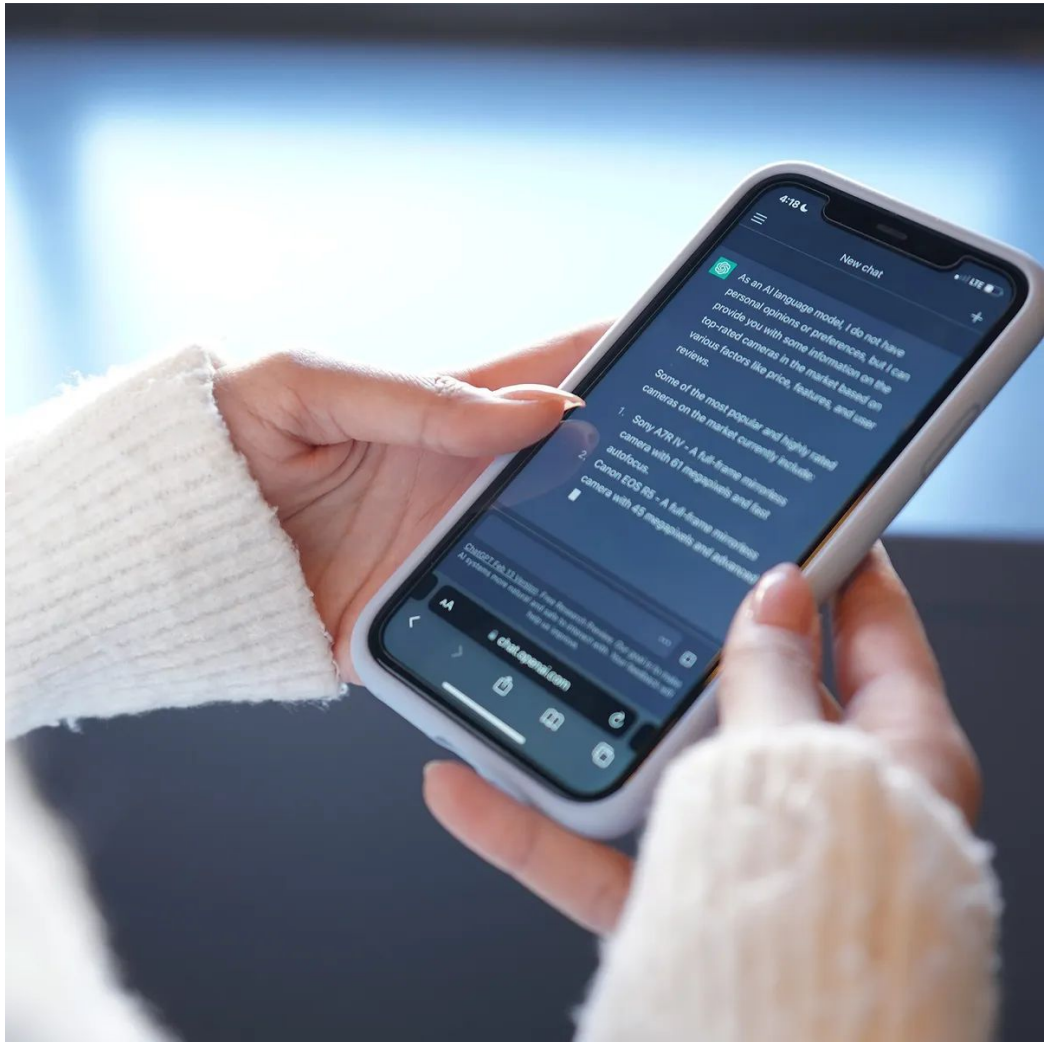
Sources:

1. https://experiments.withgoogle.com/teachable-machine

Using AI

- **Field of research:**
  - Branch of **mathematics** and **computer science**
  - Interdisciplinary by nature

- **Since 1950s** – increase in computing power allows us to achieve remarkable things!
  - **Example:** Generative AI
    - *ChatGPT, DALL·E, …*
    - *But this is not the only AI research out there!*

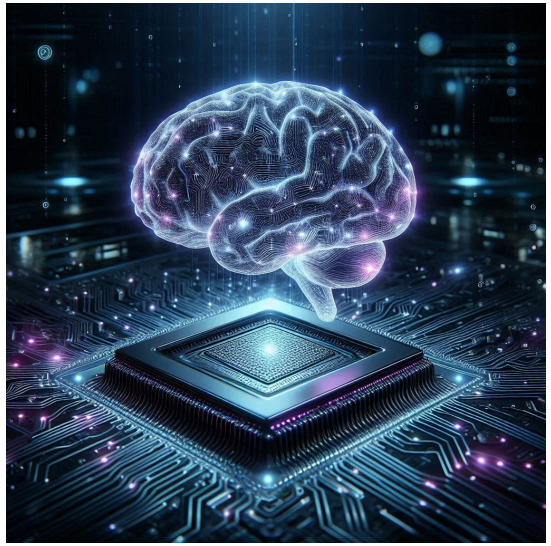


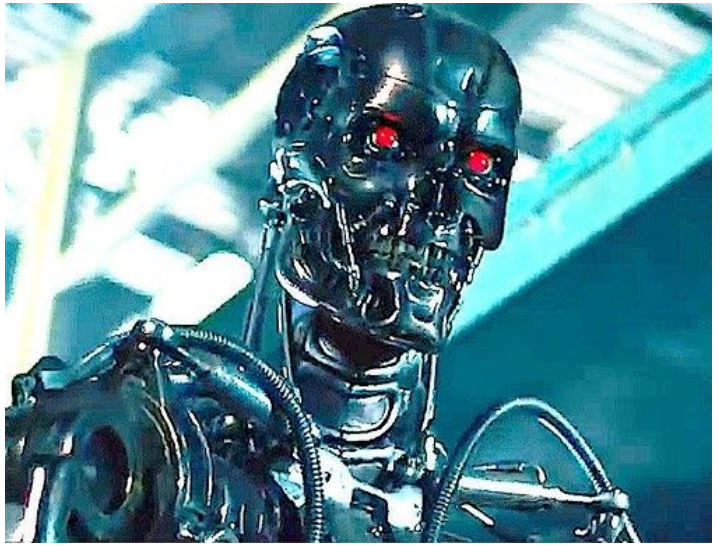
Sources:

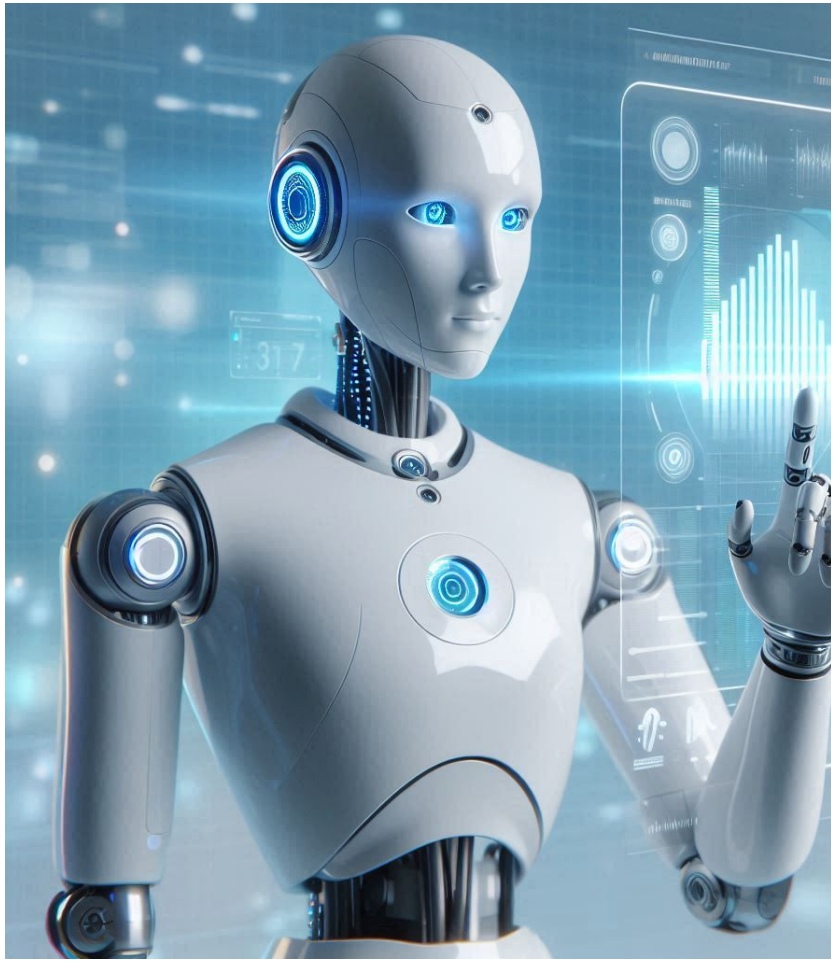
1. https://www.zdnet.com/article/chatgpt-can-now-access-the-internet-on-mobile-heres-how/

*Ab*using AI

- **Lifting common AI misconceptions:**
  - Not about conscious machines
    - AGI is a goal for some, but most research focuses on “narrow” tasks
  - *We use “an AI” to … =* 🚩🚩🚩







Sources:
1. https://www.imdb.com/title/tt0088247/

Abusing AI

Smartschool speurt via AI naar leerproblemen bij scholieren: “Het gaat niet gewoon om vinkjes tellen”

© Fred Debrock

Tom Le Bacq

Maandag 21 oktober 2024 om 09:52

“Smartschool allows parents to directly contact teachers, various school services, and the school administration, as well as check the school calendar, view reports, and monitor absences.”

Sources:

1. https://www.standaard.be/cnt/dmf20241021_92834830
2. https://www.standaard.be/cnt/dmf20241025_96029549
3. https://www.standaard.be/cnt/dmf20241024_96908130

Abusing AI

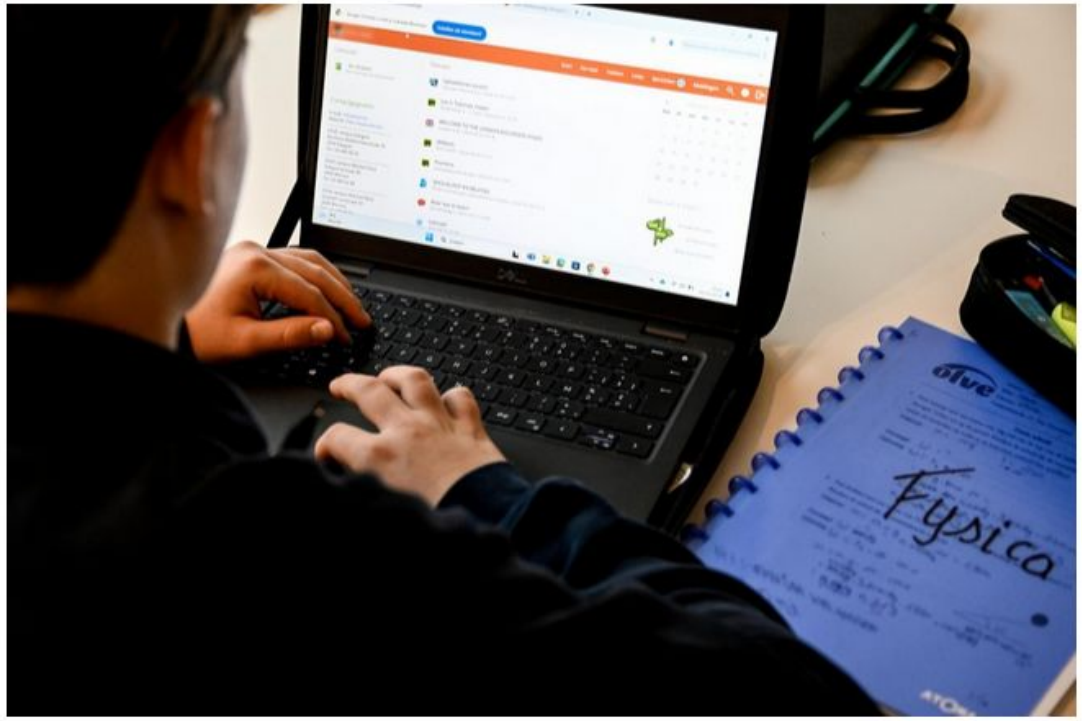
Smartschool speurt via AI naar leerproblemen bij scholieren: “Het gaat niet gewoon om vinkjes tellen”

© Fred Debrock

Tom Le Bacq

Maandag 21 oktober 2024 om 09:52

Ouders willen geen AI in Smartschool, en dat is een belangrijk signaal



© belga

Dominique Deckmyn

Zaterdag 26 oktober 2024 om 03:00

Sources:

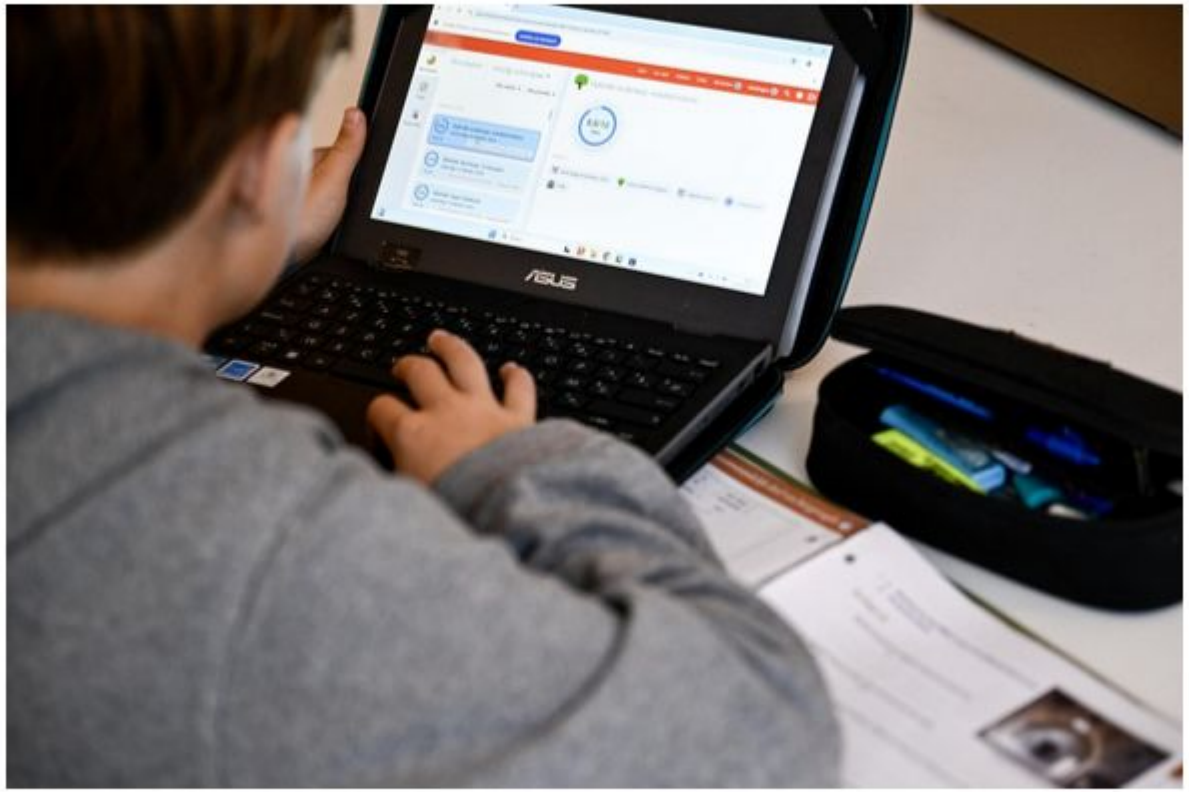
1. https://www.standaard.be/cnt/dmf20241021_92834830
2. https://www.standaard.be/cnt/dmf20241025_96029549
3. https://www.standaard.be/cnt/dmf20241024_96908130

Abusing AI

“AI will not solve the problem of students leaving education early”

OPINIE SMARTSCHOOL EN AI

AI zal leerproblemen en schoolmoeheid niet oplossen



Alsof leerlingen nog niet genoeg stress hebben — © belga

Met AI gedetailleerde leerlingenprofielen maken zonder waarborgen, schendt de rechten van jongeren, schrijft Els Kindt.

Sources:

1. https://www.standaard.be/cnt/dmf20241021_92834830
2. https://www.standaard.be/cnt/dmf20241025_96029549
3. https://www.standaard.be/cnt/dmf20241024_96908130

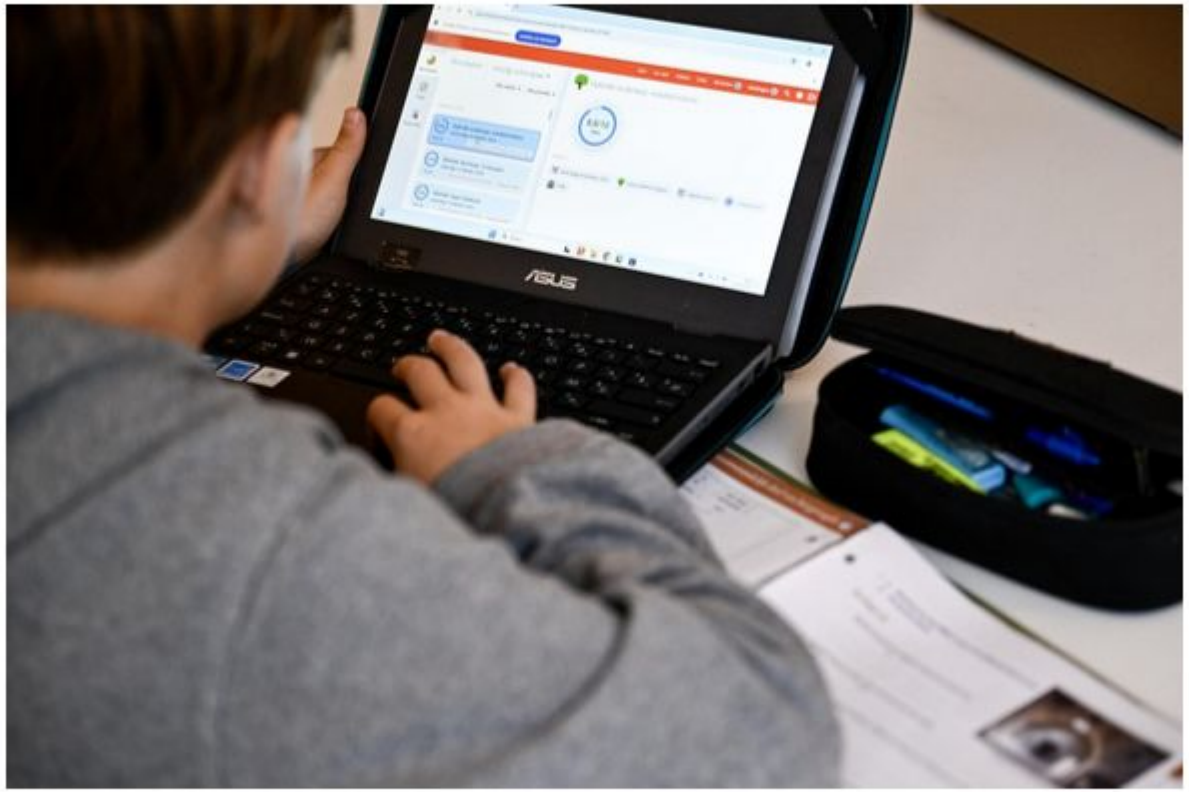
*Ab*using AI

“AI will not solve the problem of students leaving education early”

\+ Mind the ***black box***!

OPINIE SMARTSCHOOL EN AI

AI zal leerproblemen en schoolmoeheid niet oplossen



Alsof leerlingen nog niet genoeg stress hebben — © belga

Met AI gedetailleerde leerlingenprofielen maken zonder waarborgen, schendt de rechten van jongeren, schrijft Els Kindt.

Sources:

1. https://www.standaard.be/cnt/dmf20241021_92834830
2. https://www.standaard.be/cnt/dmf20241025_96029549
3. https://www.standaard.be/cnt/dmf20241024_96908130

Abusing AI

MOGELIJK VERBOD IN DE VS

TikTok aangeklaagd na zelfmoord meisjes: ‘Kwetsbare jongeren krijgen constant berichten te zien waarin suïcide als romantisch wordt voorgesteld’

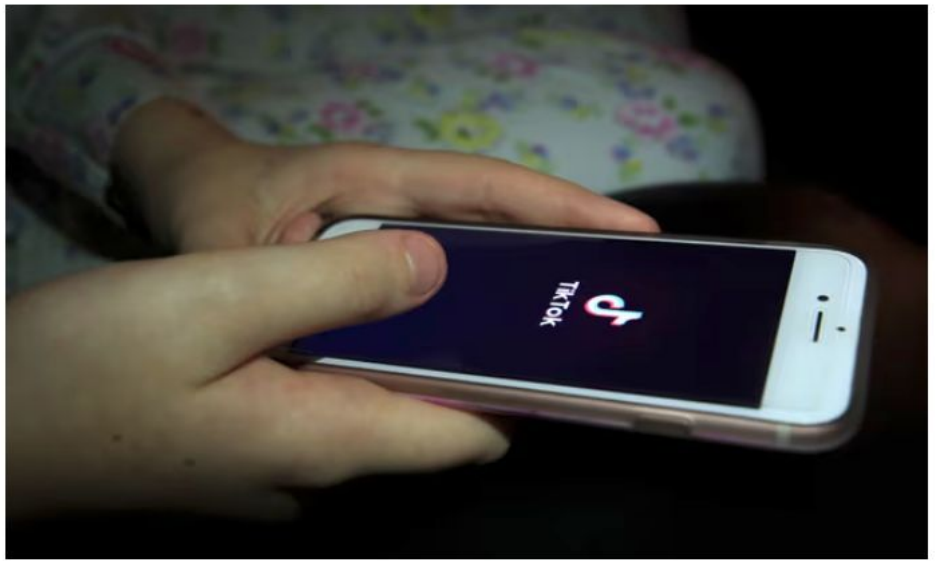
the facebook files

Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show

Its own in-depth research shows a significant teen mental-health issue that Facebook plays down in public

Social media algorithms ‘amplifying misogynistic content’

Researchers say extreme content being pushed on young people and becoming normalised



Researchers said they detected a four-fold increase in the level of misogynistic content suggested by TikTok over a five-day period of monitoring. Photograph: Peter Byrne/PA

Sources:

1. https://www.theguardian.com/media/2024/feb/06/social-media-algorithms-amplifying-misogynistic-content
2. https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739
3. https://www.humo.be/nieuws/tiktok-aangeklaagd-na-zelfmoord-meisjes-kwetsbare-jongeren-krijgen-constant-berichten-te-zien-waarin-suicide-als-romantisch-wordt-voorgesteld~b96cde5f9/

*Ab*using AI

The **Dutch childcare benefits scandal** (Dutch: *kinderopvangtoeslagaffaire* or *toeslagenaffaire*, lit. '[childcare] benefits affair') refers to a political scandal in the Netherlands involving false allegations of welfare fraud by the Tax and Customs Administration (*Belastingdienst*) against thousands of families claiming childcare benefits.[1][2]

Between 2005 and 2019, approximately 26,000 parents were wrongly accused of making fraudulent benefit claims, resulting in demands to repay their received allowances in full.[1][3] In many cases, this sum amounted to tens of thousands of euros, driving families into severe financial hardship.[1][2]

Sources:
1. https://en.wikipedia.org/wiki/Dutch_childcare_benefits_scandal
2. https://3bplus.nl/de-rol-van-artificial-intelligence-algoritmes-bij-de-toeslagenaffaire/

*Ab*using AI

The **Dutch childcare benefits scandal** (Dutch: *kinderopvangtoeslagaffaire* or *toeslagenaffaire*, lit. '[childcare] benefits affair') refers to a political scandal in the Netherlands involving false allegations of welfare fraud by the Tax and Customs Administration (*Belastingdienst*) against thousands of families claiming childcare benefits.[1][2]

Between 2005 and 2019 making fraudulent benef allowances in full.[1][3] In euros, driving families in

Geautomatiseerd risicoselectiesysteem

De Belastingdienst werkte in het geval van de toeslagenaffaire met een geautomatiseerd risicoselectiesysteem op basis van Artificial Intelligence (AI), dat bepaalde welke toeslagaanvragen extra gecontroleerd moesten worden. Hierbij was 'dubbele nationaliteit' bijvoorbeeld een van de selectiecriteria. Het resultaat was dat toeslagaanvragers met een tweede nationaliteit meer kans liepen om eruit gepikt te worden door het AI-algoritme.

\+ Mind the ***black box***!

Sources:
1. https://en.wikipedia.org/wiki/Dutch_childcare_benefits_scandal
2. https://3bplus.nl/de-rol-van-artificial-intelligence-algoritmes-bij-de-toeslagenaffaire/

1. **Using AI** to detect the unusual
2. Using AI to **detect the unusual**
3. **Using AI to detect the unusual**

Data?

- **Gathering data:**
  - *“Every process, system, or activity around us can produce data if we choose to observe and measure it.”*

- **Examples:**

Health Monitoring

- Heartbeats
- Blood pressure

Financial Activity

- Transaction date, time, location, and amount
- Store purchases, cash withdrawals, transfers

Social Network Activity

- Messages
- Friend requests
- Likes and comments

Anomalies?

- **Deviations from what we consider “normal”:**
  - These deviations **leave a trace in our observations**
  - We are interested in the **underlying cause** of these deviations

- **Examples:**

Health Monitoring

- Arrhythmia (*visible on ECG*)
- Blood pressure spikes + drops (*visible in BP measurements*)

Financial Activity

- Unexpected foreign transactions
- Unusual purchases, large cash deposits or withdrawals, …

Social Network Activity

- Spam messages
- Spam friend requests

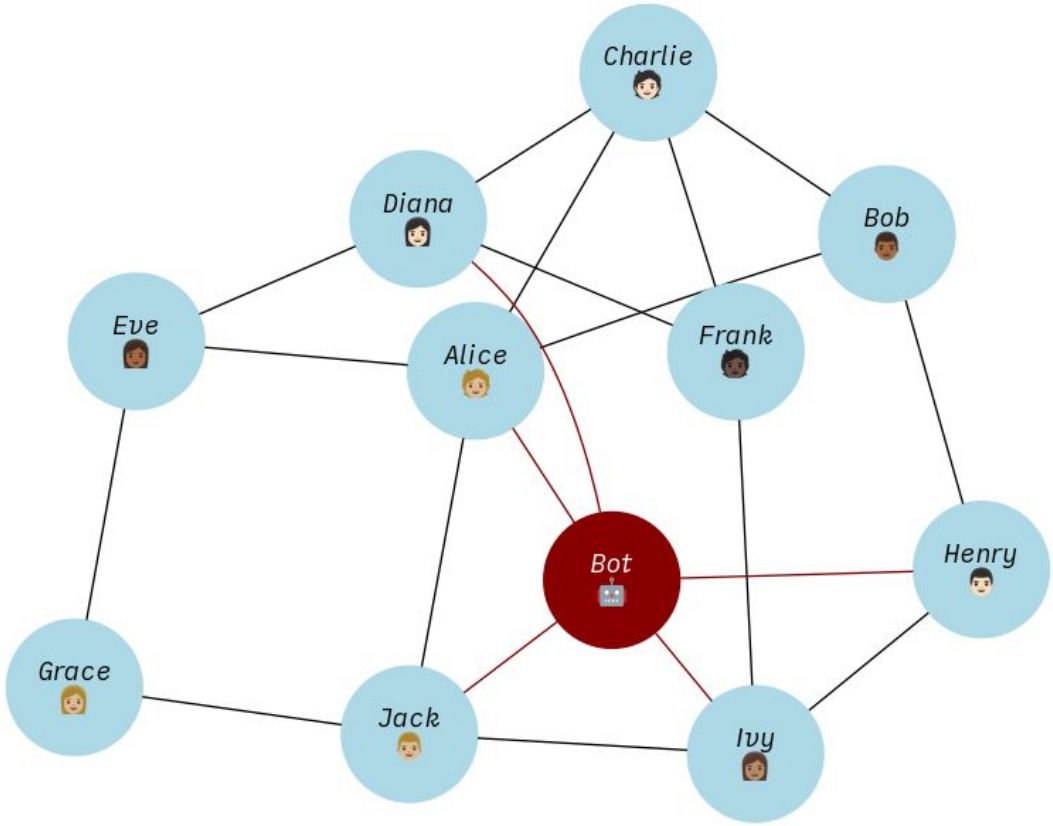
Anomalies?

- **Deviations from what we consider “normal”:**
  - These deviations **leave a trace in our observations**
  - We are interested in the **underlying cause** of these deviations

- **Examples:**

Social Network Activity

- Spam messages
- Spam friend requests

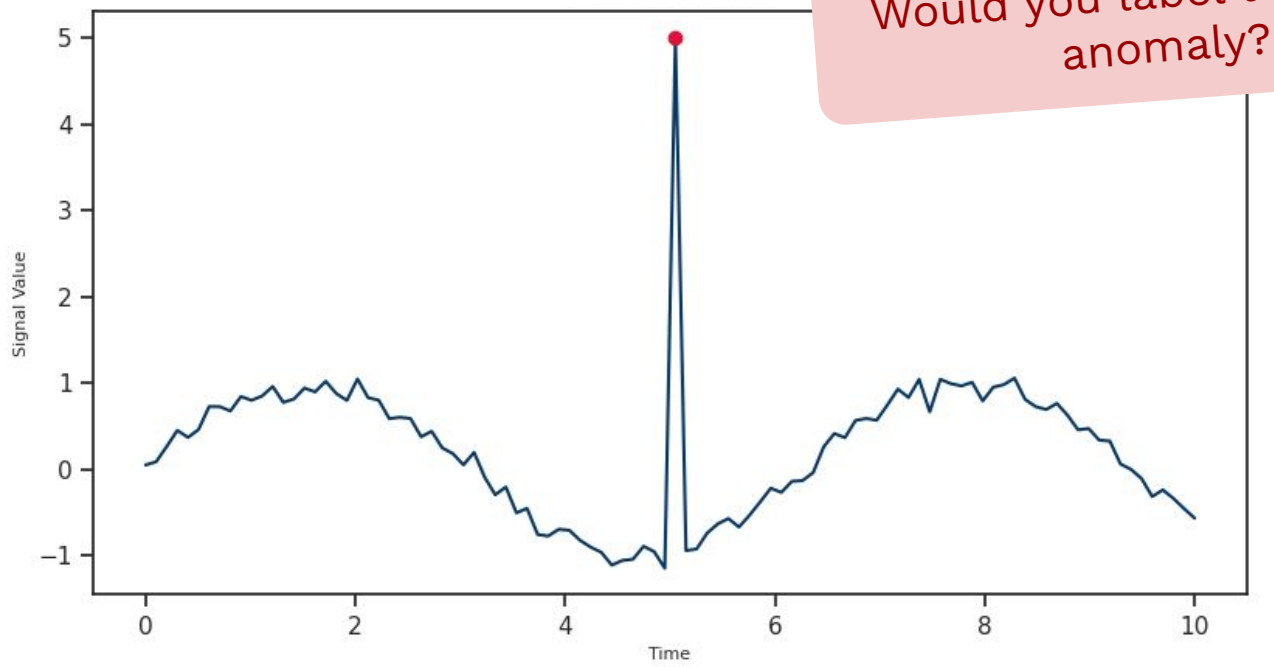


Anomalies and Outliers

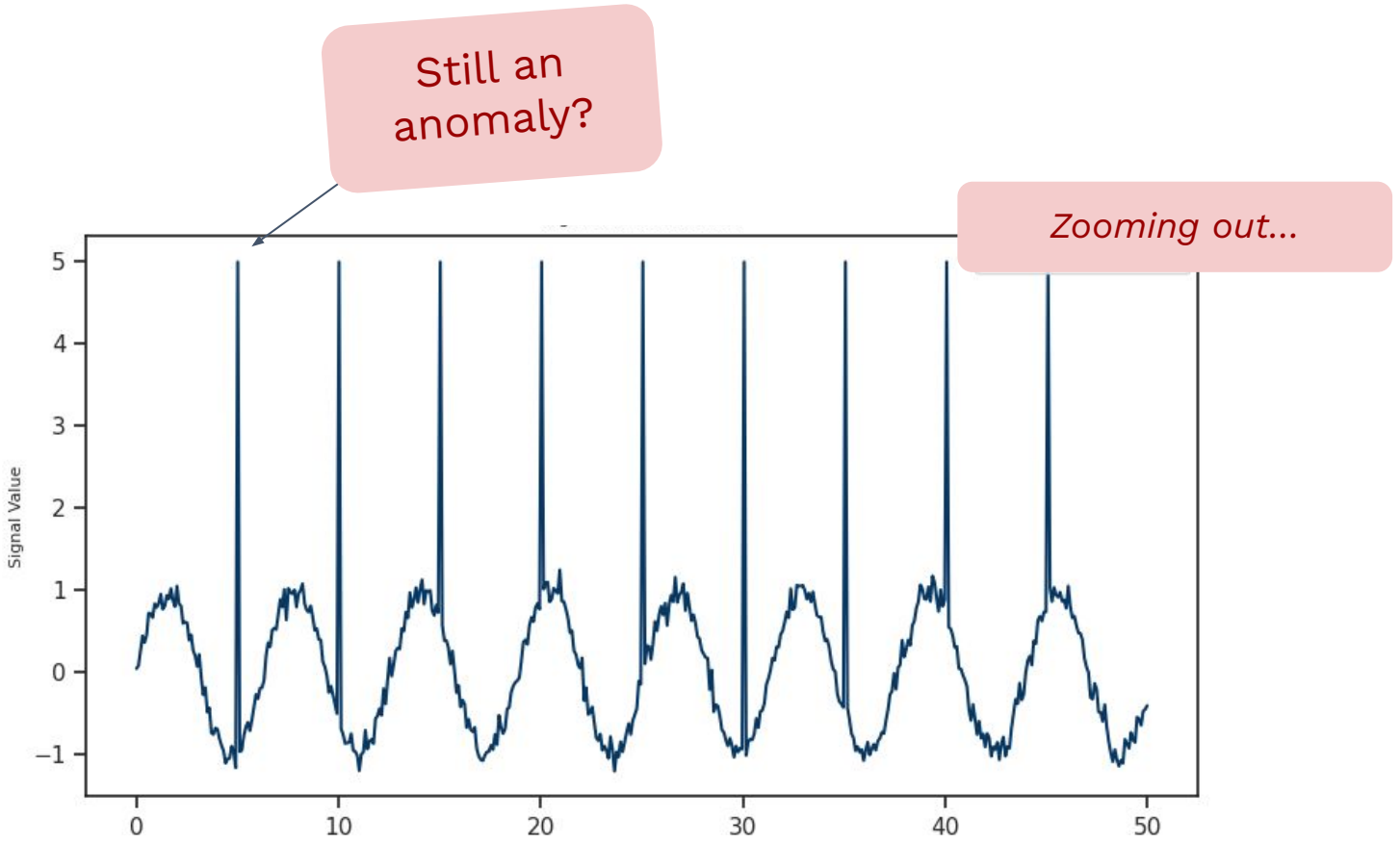
- **Some sources claim that anomaly detection = outlier detection**

- **But <u>not all outliers are anomalies</u>, and <u>not all anomalies are outliers</u>:**
  - Traditional examples mostly show
    - high peaks,
    - sudden drops, and
    - obvious outliers as anomalies
  - **But this is not always the case!**

Context Matters



Context Matters



Context Matters

- **An observation being anomalous is context-dependent:**
  - If *X* takes place before or after *Y*, it is normal / abnormal
  - If *X* takes place during *Y*, it is normal / abnormal
  - If *X* happens at regular intervals, the absence of *X* is abnormal
  - *... and so on*

- **Examples:**
  - **Heart rate:** exercise vs. rest
  - **Web traffic:** cyber-attack vs. sale
  - **Device activity:** work hours vs. downtime

1. **Using AI** to detect the unusual
2. Using AI to **detect the unusual**
3. **Using AI to detect the unusual**

Detecting Anomalies

- **Anomaly detection:**
  - *"Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behavior"* (Chandola, 2009)

- **Why detect anomalies?**
  - Alert domain experts that "something is different"
  - Preferably providing actionable information to these experts

- **Differs from classification:**
  - **Unbalanced:** anomalies are rare, so "classes" are unbalanced
  - **Open set problem:** we don't always know how certain anomalies show themselves in data

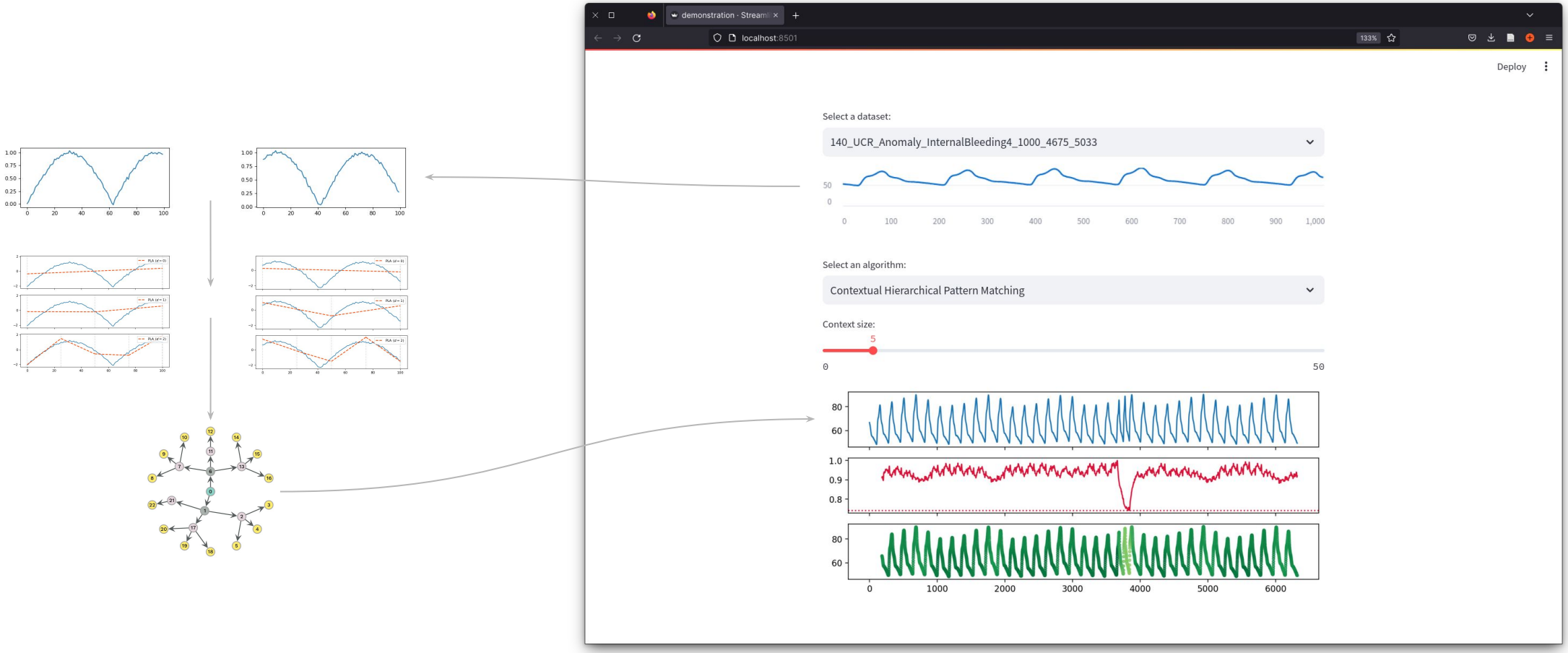
Sources:

1. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, *41*(3), 1-58.

Rule-based Approach

- **Manual rule-based configuration for anomaly detection:**
  - **Some anomalies are missed** when predefined rules do not trigger
    - No rule for certain anomalies = these anomalies go undetected
  - Rule-based trigger check on a “per event” basis, **no “bigger picture”**
    - Context matters, and anomalies can involve multiple events
  - **Expert time wasted** on hard-coded trial-and-error design

The **pattern-based series framework** (`pbsf`) is designed to help us implement and test algorithms that operate on sequential data. These algorithms typically involve segmenting data into smaller parts, discretising these parts, and then combining the results into a larger aggregate structure. This framework can be utilised for tasks such as anomaly detection and series classification.



Difficult Challenges

- **Complexity of normal behaviour**
  - Learning what is “normal” is not straightforward
  - Long-term dependencies are hard to model

- **Domain-specific definitions:** interdisciplinarity matters
  - Normal in one domain could be abnormal in the other
  - There is no one-size-fits-all detection algorithm

- **Ambiguity of artefacts:** context matters
  - Different events can show themselves similarly in data

Impact of Mistakes

- **Algorithms can get things wrong:**
  - **False positive:** things flagged as anomalous may actually be normal
  - **False negative:** things considered as normal may be anomalous

- **Trade-offs matter:**
  - **Example:** “smoke detectors, bad cooking, and housefires”

- **Going back to our (ab)use of AI:**
  - The impact of FP and FN

Impact of Mistakes

Smartschool speurt via AI naar leerproblemen bij scholieren: “Het gaat niet gewoon om vinkjes tellen”

What is the impact of a *false positive*?

What is the impact of a *false negative*?

Wat is de toeslagenaffaire?

Ouders krijgen geld van de Belastingdienst om kinderopvang voor hun kinderen te betalen. Dat heet kinderopvangtoeslag. De Belastingdienst heeft tussen 2005 en 2019 fouten gemaakt. Sommige ouders kregen te horen dat ze duizenden euro's moesten terugbetalen terwijl ze wel recht hadden op dit geld.

Key Takeaways

- **AI is a field of research:**
  - Look past the marketing hype!

- **AI and anomaly detection:**
  - Identify unusual patterns to detect anomalies in diverse domains
  - Context-dependent deviations from what we consider normal

- **Algorithms can make mistakes:** *mind the human*
  - **False positives:** normal events flagged as anomalies
  - **False negatives:** anomalies flagged as normal
  - **Consequences of these errors may be disastrous!**

Slides and transcript can be found at:

pvdsp.be/posts/fari-2024