ComputationWorld 2023 & NexComm 2023

Theme

Quality of Data/Information and the Accuracy AI-based Tools

- From "Garbage IN, Garbage OUT" to Cleansing and Deep Learning

- Human Common-sense decisions vs. Machine fact-based decisions in driving and possible discrepancies

- Awareness on Data accuracy and Precaution on Trusting Information Processing

- How to Securely and Safely benefit from AI-based technologies advances with no harms for Humans?

- Case studies (avionics, automotive, agriculture, etc.)

Mgr. Dr. **Mirnig Alexander**, AIT Austrian Institute of Technology,

Center for Technology Experience, Austria, IARIA Fellow

Prof. Dr. **Petre Dini**, *IARIA, USA/EU, IARIA Fellow*

- Knowledge and Information of any kind
  ----> ~ 99% undetected/not check-able false information
  (on purpose coalitions, or simple fake, or science fiction)
  and
  -----> ~1% fact-checked information
  ---- >>> any inference leads to biased, unproved, and
  unanimously (falsely) trusted facts/conclusions

  My conclusion: no intelligence (artificial, human) can be
  trusted until 'fact checked' and properly interpreted.
  Facts checking: validation, logic, transparent process, etc.
  Conclusion: Human common sense prevails

BIG | the Vs | 3v, 5v, 7v, 10v, …. ?

- **Volume** (length of a records, # of records) (entity-relationship databases)(datasets) || **BIG vs. HUGE**

- **Variety** (types: strings, pictures, voice, etc.) (structured, non-structured)

- **Veracity**   (**precision** and **accuracy** of data)

- **Velocity** (of change)

- **Value** (as a business/service) **IMPACT**

- **Volatility** (temporary; quick action)

- **Vasting resources**
  (storage, computation, transfer)

- **Viability** (are data still useful?)

- **Visibility** (open, hidden, ..)

- **Validity**
  (are there still valid/updated data?)
  (in context validity)
  (e-government datasets)

- incomplete
- redundant
- inconsistent
- noisy

quality of data

filling missing values with estimated values calculated for complete records of the same dataset

- **AI-based Health Applications**
  - Pharmacology
  - Personal healthcare
  - Medical accuracy and precision
  - Medical Robots

- **Society domains**
  Automotive
  Avionics
  Industry (Industry 5.0)
  Agriculture (Precision agriculture)
  Financial (Gain Optimization)
  Energy (Solar Panel/Windmill management)
  etc.

- **Data Processing is a complex process**
  - Profiling (statistical analysis, pattern recognition, and data visualization)
  - Cleansing (deduplication, standardization - consistent formats and units-, validation (against predefined rules or reference data, eyc.)
  - Validation (data type validation, range validation, format validation, and referential integrity checks to ensure data consistency)
  - Establishing data quality metrics (completeness, accuracy (error/correctness), consistency (e.g., percentage of conflicting data), timeliness (e.g., data freshness or latency), and uniqueness (e.g., the number of duplicate records).)

Awareness on Data Provenance and Data Pre-processing
Explainability of processes and output

CASE STUDY: Mobility

Mobility

- Automated mobility is reliant on sensing accuracy
  - Geographical Positioning
  - Distance to other objects
  - Object recognition
  - Trajectory estimation (own and other agents')

- Reduction in accuracy leads to
  - Decreased performance, safety, and mitigating measures (e.g., lower overall speeds), eventually resulting in
  - A reduced Operational Design Domain (ODD)

Recommender Systems

- Provide condensed, action-specific output, often on the basis of large volumes of data that are too big to be easily processed by humans.
    - E.g., process management systems

- Modern recommender systems are frequently AI-based to increase robustness and enable extension towards process modifications, additional parameters or agents (e.g., machine types).

- Recommender systems typically do not evaluate their output in-UI (if below threshold, it will not be chosen or shown).

What if…

- Automation can be seen analogous to recommender systems

- A recommender, instead of providing single-level output, can provide *recommendations on its recommendations*.

- The output is accompanied by information on the expected reliability of the output

- Not all data sources are equal, output should reflect that.

- AI can estimate and inform about their own estimated accuracy, thusly enable more informed use of the system

- Thereby: anticipate and prevent errors or undesired consequences, integrate the human in the loop as an informed agent, extend the scope (or ODD) of the technology.

Stage for the Audience