# AI in Governance: Making it work


## Masood Mehmood Khan

**Mechatronic Engineering**
**and**
**SmartCrete@Curtin**


**23rd Symposium on Artificial Intelligence**
**Pakistan Academy of Engineering, Karachi Pakistan**

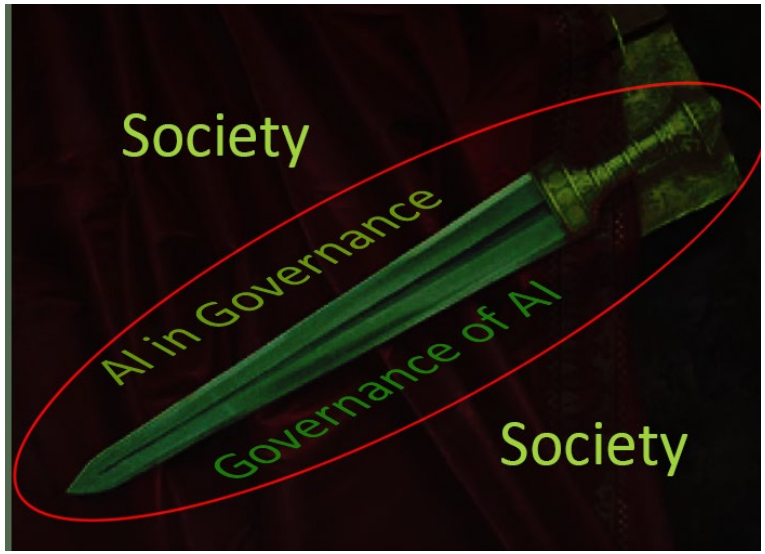
26 March 2022



Curtin University

# AI and human society

- Is this a dagger in the hands of an assassinator?
- Could this be a precision knife in the hands of a surgeon?
- How should a society (like ours) interact with AI?
- Which one of the two edges is absolutely benign?

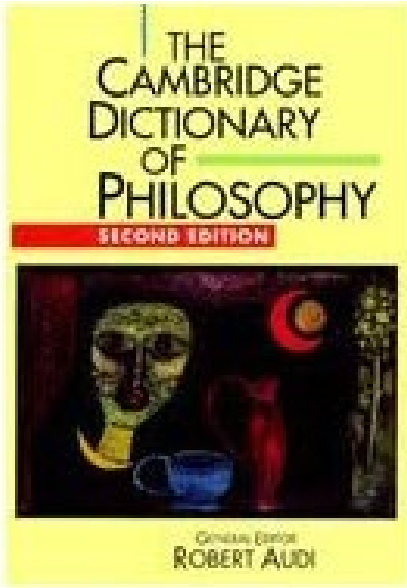


# AI in Governance

- Knowing the risks of falling on any of the two edges, let's focus on AI in Governance.

# Governance

- The Cambridge Dictionary of Philosophy does not provide a definition of governance.
- The nearest term in there is Casuistry - the case analysis approach to the interpretation of general moral rules.
- Casuistry looks into paradigm cases of how and when a given moral rule would be relevant.
- Casuistry reasons by analogy to cases in which the proper application of the rule is less obvious, e.g., a case in which lying is the only way for a priest not to betray and not to reveal a secret told in confession.



# Governance

- **Casuistry** is the use of clever but unsound reasoning, especially in relation to moral questions.
- **Casuistry** is the process of resolving moral problems by the application of theoretical rules.
- The term **Governance** encompasses the system by which an organisation is controlled and operated, and the mechanisms by which it, and its people, are held to account [Governance Institute].
- **Governance** refers to all the processes of interactions; through the laws, norms, power or language of an organized society over a social system (family, tribe, formal or informal organization, a territory or across territories). It is done by the **government** of a state, by a **market**, or by a **network** [Wikipedia].

# Governance

- AI systems now make critical decisions that affect society and contribute to governance.
- One of the most important research problem is to understand the ethical implications of AI systems' decisions and actions.
- Developing such an understanding would require integrating moral, legal, societal and historical considerations in the AI system/s.
- Hence, the AI system design process and the real-life assessment of system decisions need to be envisaged, reviewed and recommended.

# AI in Governance

- COMPAS, a system that assesses offenders' criminogenic needs and risks of recidivism.

  Instigated legal debate and deliberations on accountability and lack of transparency in AI systems (Liu, Lin & Chen, 2019).

*International Journal of Law and Information Technology*, 2019, 27, 122–141
doi: 10.1093/ijlit/eaz001
Advance Access Publication Date: 12 February 2019
Article

OXFORD

## Beyond *State v Loomis*: artificial intelligence, government algorithmization and accountability

Han-Wei Liu,* Ching-Fu Lin† and Yu-Jie Chen‡

ABSTRACT



Curtin University

# AI in Governance

- Biometrics and biometric-enabled decision-support systems (DSS) are a mandatory part of complex dynamic systems.
  - Used at security checkpoints, in personal health monitoring systems, autonomous robots, and epidemiological surveillance.

SPECIAL SECTION ON INTELLIGENT BIOMETRIC SYSTEMS FOR SECURE SOCIETIES

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 21, 2020, accepted August 3, 2020, date of publication August 11, 2020, date of current version August 24, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3015855*

## Risk, Trust, and Bias: Causal Regulators of Biometric-Enabled Decision Support

KENNETH LAI[1], HELDER C. R. OLIVEIRA[1], MING HOU[2], (Senior Member, IEEE), SVETLANA N. YANUSHKEVICH[1], (Senior Member, IEEE), AND VLAD P. SHMERKO[1]

[1]Biometric Technologies Laboratory, Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada
[2]Defence Research and Development Canada (DRDC), Ottawa, ON K1N 1J8, Canada

Corresponding author: Kenneth Lai (kelai@ucalgary.ca)

Curt
CRI

rtin University

# Summary of problems


- Algorithmic issues

  Accountability, Explainability, Fairness, Interpretability, Robustness, Security, Transparency

- Public policy issues

  Competition, Control, Legal Framework, Ownership, Policy Influence





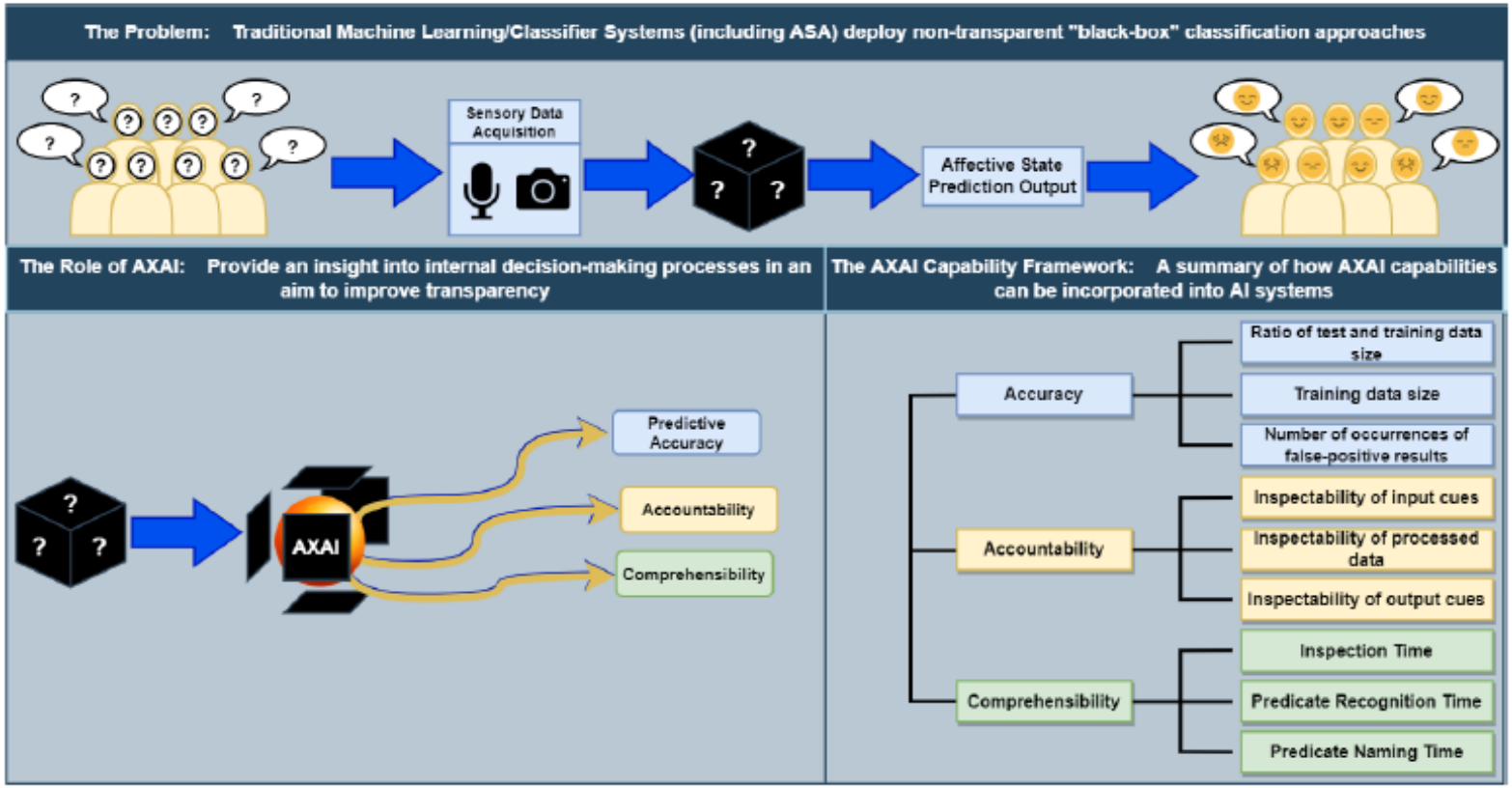
Curtin University

# Create a testbed


- Establish theoretical foundations
- Avoid near-accurate single-cue models to improve system accuracy.
- Exploit the holism of a multimodal solution
- Build a prototype, stand-alone, real-time multimodal affective state classification system.
- Use speech and facial muscle movements to create a classifier.
- Validation accuracies of 99.2% and 92.78% achieved.

# The anticipated change


- Enable design and implementation of acceptable AI systems




University

# The testbed

- Establish theoretical foundations
- Validation accuracies of 99.2% and 92.78% achieved.


Raw Input Data Screen of the A-Machine software. The image is an example of extracting expression of anger - the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS).

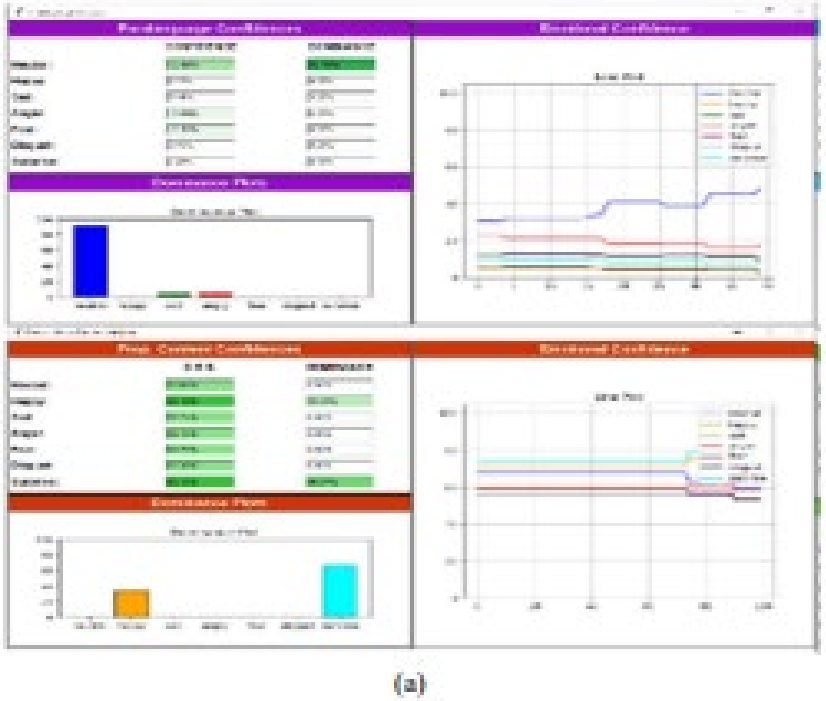




Curtin University

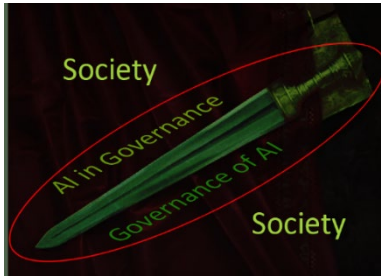
# The testbed

The two images show the left (a) and right (b) hand sides of a GUI screens for Paralanguage (top-left of the screen) shown in (a). Facial Affect (top-right of the screen) shown in (b). Propositional Content (bottom-left) shown in (a). The final Rule-based Expert System (bottom-right of the screen) shown in (b). Each window highlighted in this figure is viewed simultaneously vis-à-vis viewing the input data screen shown on previous slide.



(a)

# Where would the testbed fail?



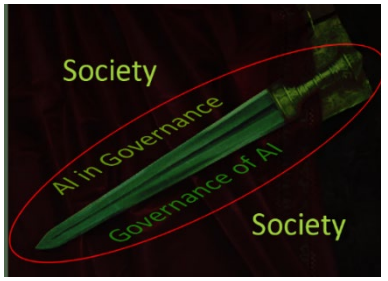

- Features leading to decisions were displayed but explicit explanations were missing.
- Pieces of information appear to be convoluted in the displayed implicit explanations: inferences and outcomes were not clear.
- Establishing accountability for the impact of outcomes was not possible.
- A chain of responsibility for either success or failure of the system could not be established.





Curtin University

# Developing theoretical foundations




- Propose an Accountable eXplainable Artificial Intelligence (AXAI) capability framework.
- Separate and quantify elements of comprehensibility, accuracy and accountability - facilitate incorporating and assessing XAI capabilities.
- Enable delineation of AI systems in a three-dimensional (AXAI) space.

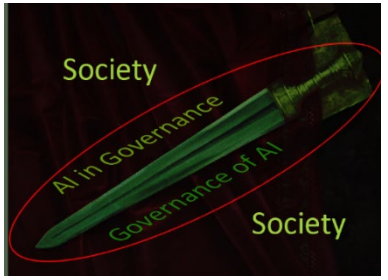
**Toward Accountable and Explainable Artificial Intelligence Part one: Theory and Examples**

Cite | Download (1.14 MB) | Share | Embed

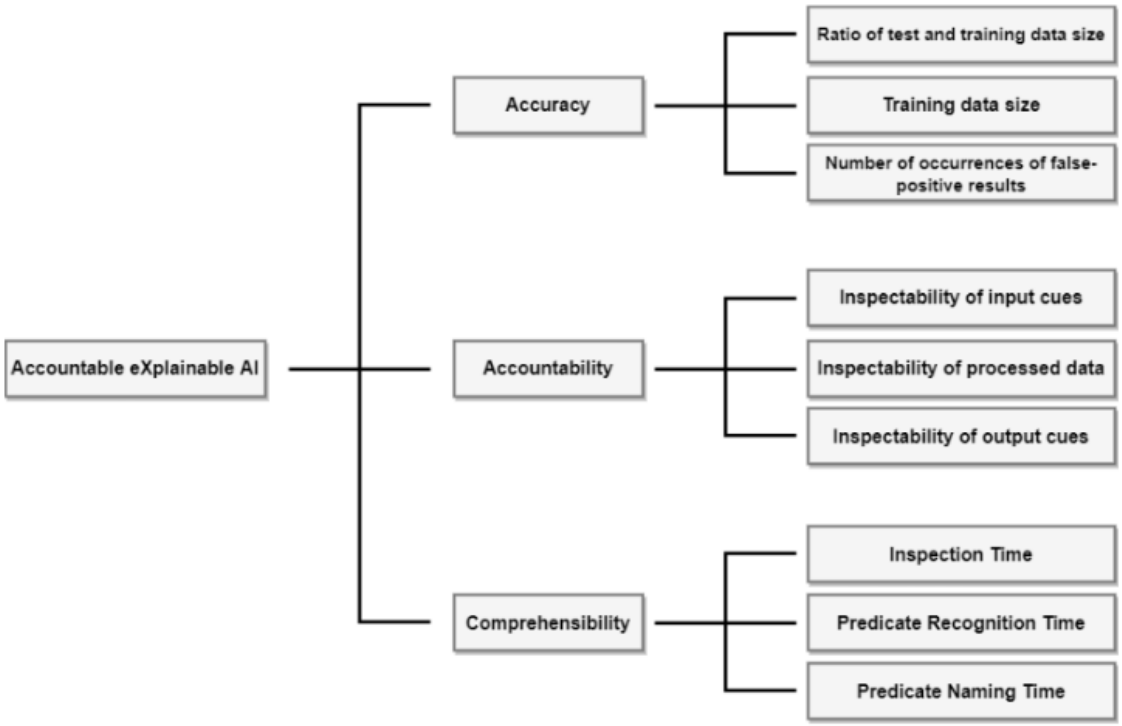
Preprint posted on 10.02.2022, 20:34 by Masood Khan, Jordan Vice

USAGE METRICS

# Developing the 3D AXAI space




The three-level Galois-lattices structure representing the proposed explainability.

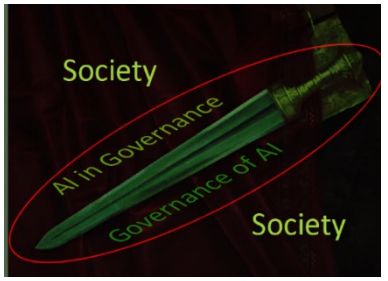







Curtin University

# Developing the framework




## Inside the AXAI framework

### 3.2 Definitions

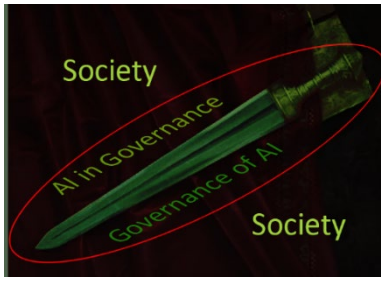
***D1***: A predicate symbol, usually called in queries, is such that $p \in \mathcal{P}$. Declared in an AI system ($\mathcal{P}$) is $p$, which is *public* with respect to a human population $\mathbb{S}$ if $p$ forms part of the background knowledge $\mathbb{B}$ of each human $s$ ($s \in \mathbb{S}$). Otherwise, $p$ is a *private* predicate symbol contained in $\mathcal{P}$.

***D2***: Let $\mathbb{H}$ be some AI system. If the background knowledge $\mathbb{B}$ of $\mathcal{P}$ is extended such that $\mathbb{B} \cup \mathbb{H}$ is formed, then the predicate symbol $p \in \mathcal{P}$ becomes a predicate invention since $p$ was originally defined in $\mathbb{H}$ but not in $\mathbb{B}$.

***D3***: The Accountable Explainable Artificial Intelligence (AXAI) capability denoted by $X_{AAI}$ is a representation in a three-dimensional space. We posit that $X_{AAI}$ comprises of three independent vectors: $\mathbb{C}$ (comprehensibility), $P_A$ (predictive accuracy) and $S_A$ (system accountability). Also, each one of the three vectors $\mathbb{C}$, $P_A$ and $S_A$ comprises of three independent components whose details are given in the following definitions ***D4 – D6C***.

***D4***: The comprehensibility $\mathbb{C}$ of $\mathcal{P}$ in the context of a human population $\mathbb{S}$ is represented as $\mathbb{C}(\mathbb{S}, \mathcal{P})$ where $\mathbb{C}$ is a vector comprising of three components: the inspection time ($T_{it}$), the predicate recognition time ($T_{pr}$) and the time required to name a predicate ($T_{pn}$) such that:

$$\mathbb{C}(\mathbb{S}, \mathcal{P}) = \sqrt{(T_{it}^2 + T_{pr}^2 + T_{pn}^2)} \quad (1)$$




# Implementing the framework

## The revised system

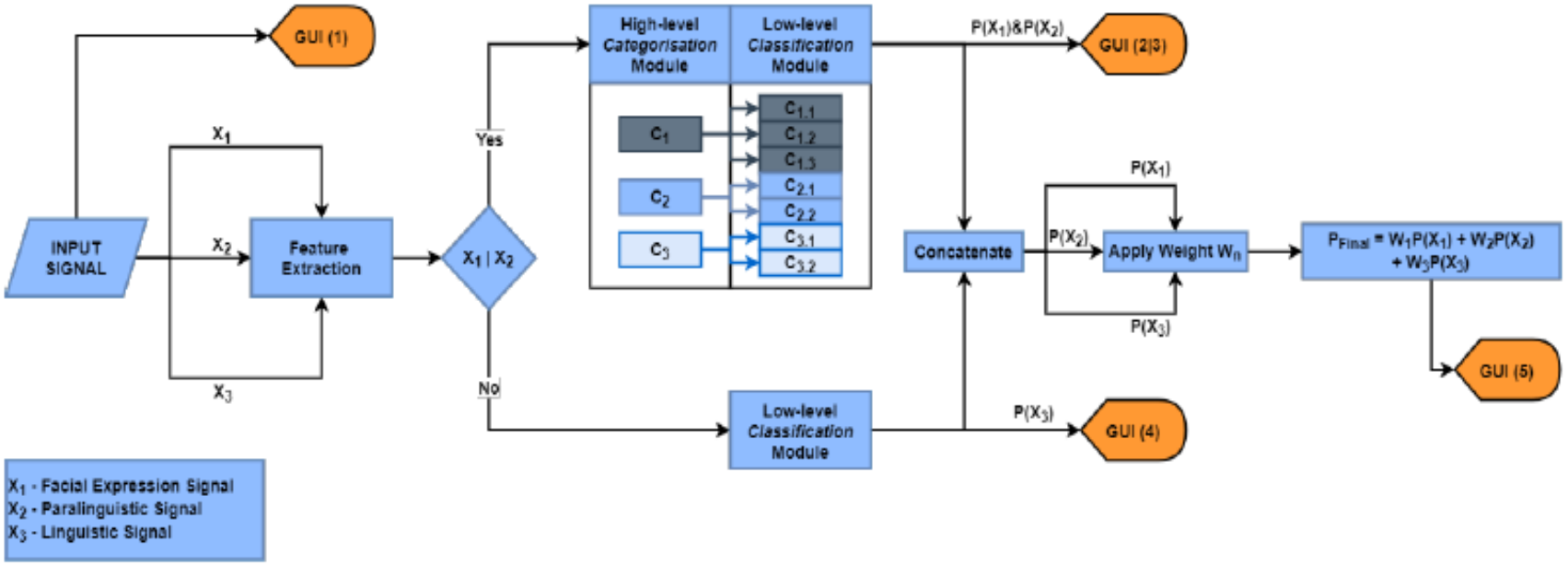


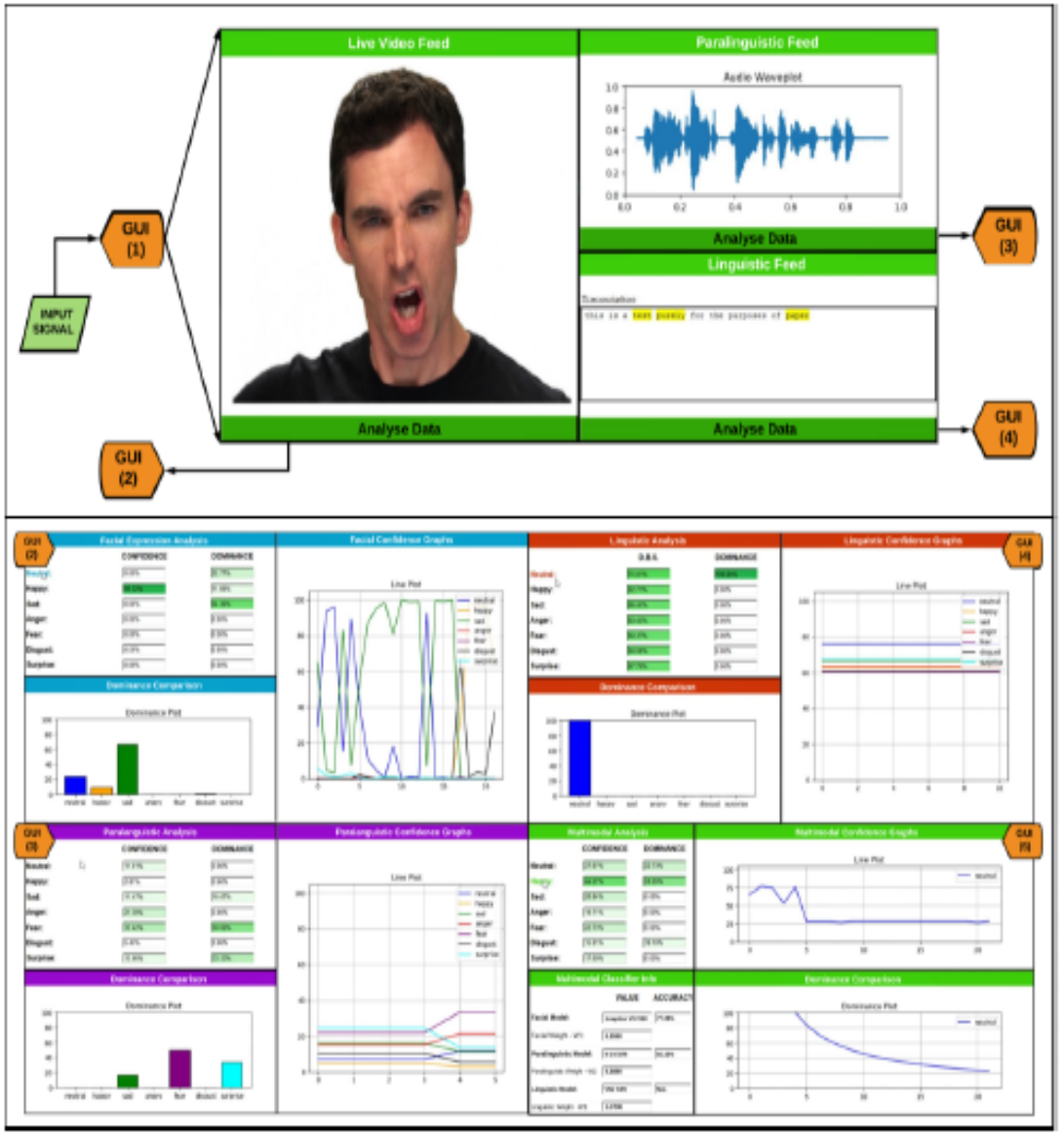

Figure 3: ASAM high-level system architecture, highlighting how signals $\{X_1, X_2, X_3\}$ traverse throughout the system through processing and output stages and how/what information is displayed to the user via the graphical user interface (GUI) shown in Figure 4

# Implementing the framework

ASAM GUI showing the input, processing and output info. The upper window is shown to users upon execution of the software. The shown image is an example from the RAVDESS dataset.

The four coloured windows in the lower half of the figure are the monitoring windows used to investigate processing and output information. CYAN = Facial Expression Analysis, ORANGE = Linguistic Analysis, PURPLE

= Paralinguistic Analysis and LIME = Multimodal Analysis. All windows are executed on separate threads allowing simultaneous viewing.



# AXAI capable software design

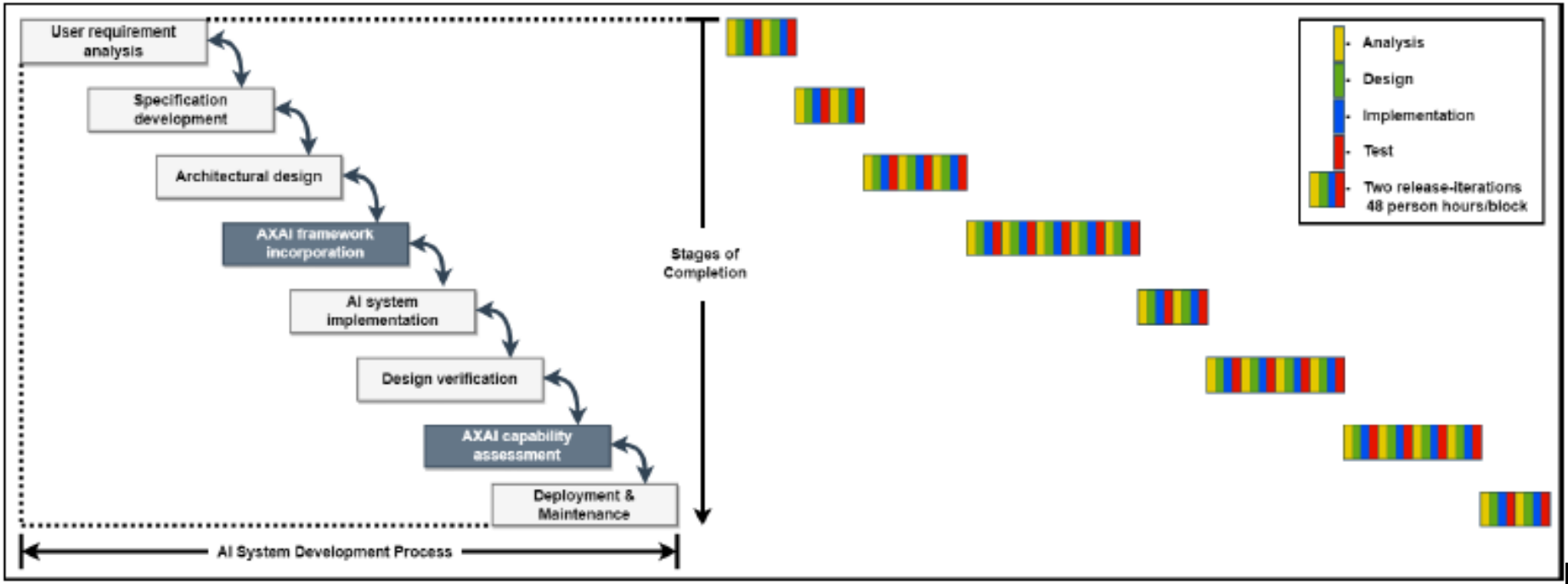
- Demonstrate incorporating the AXAI capability in a robotic system while adhering to the eXtreme Programming (XP) practices.
- Enable understanding user behaviour while systematically incorporating the AXAI capability. Identification of ethical, technical, functional, and domain-specific system requirements are resolved using an appropriate software design process.

**Toward Accountable and Explainable Arti icial Intelligence Part Two: The Framework Implementation**

JORDAN VICE[1] AND MASOOD M. KHAN[1]

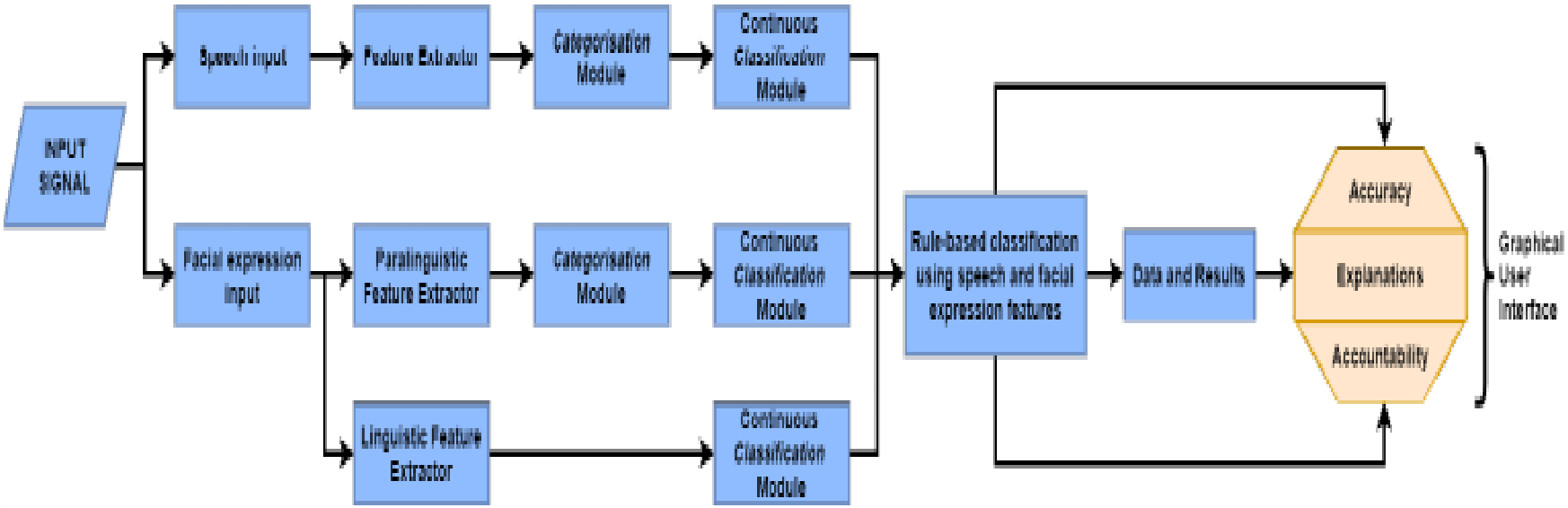
# AXAI capable software design process

- We could map the conventional AI system design and development process to the extreme programming practices. Each block of yellow, green, blue and red boxes represents a set of two release-iterations.



# AXAI capable software design process

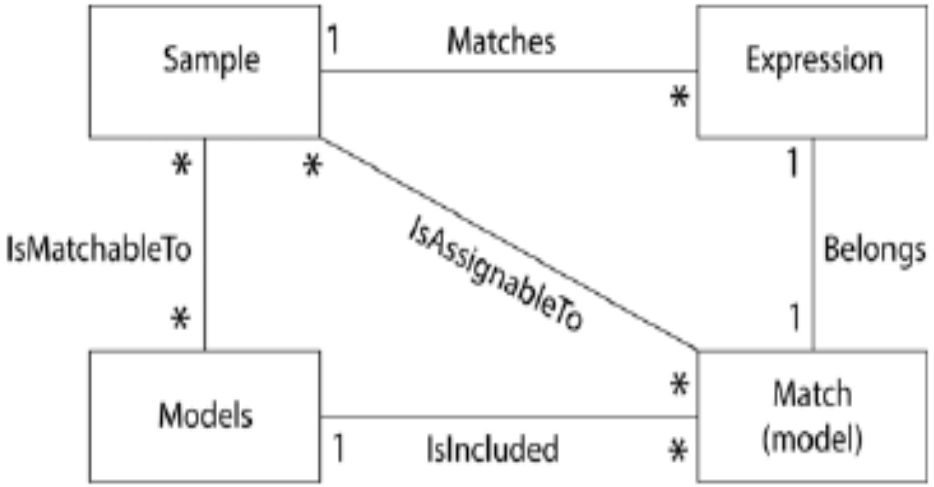
- Functional modules and their interconnections within the ASAM. All elements of accuracy, comprehensibility and accountability are embedded into the graphical user interface.



# AXAI capable software design process

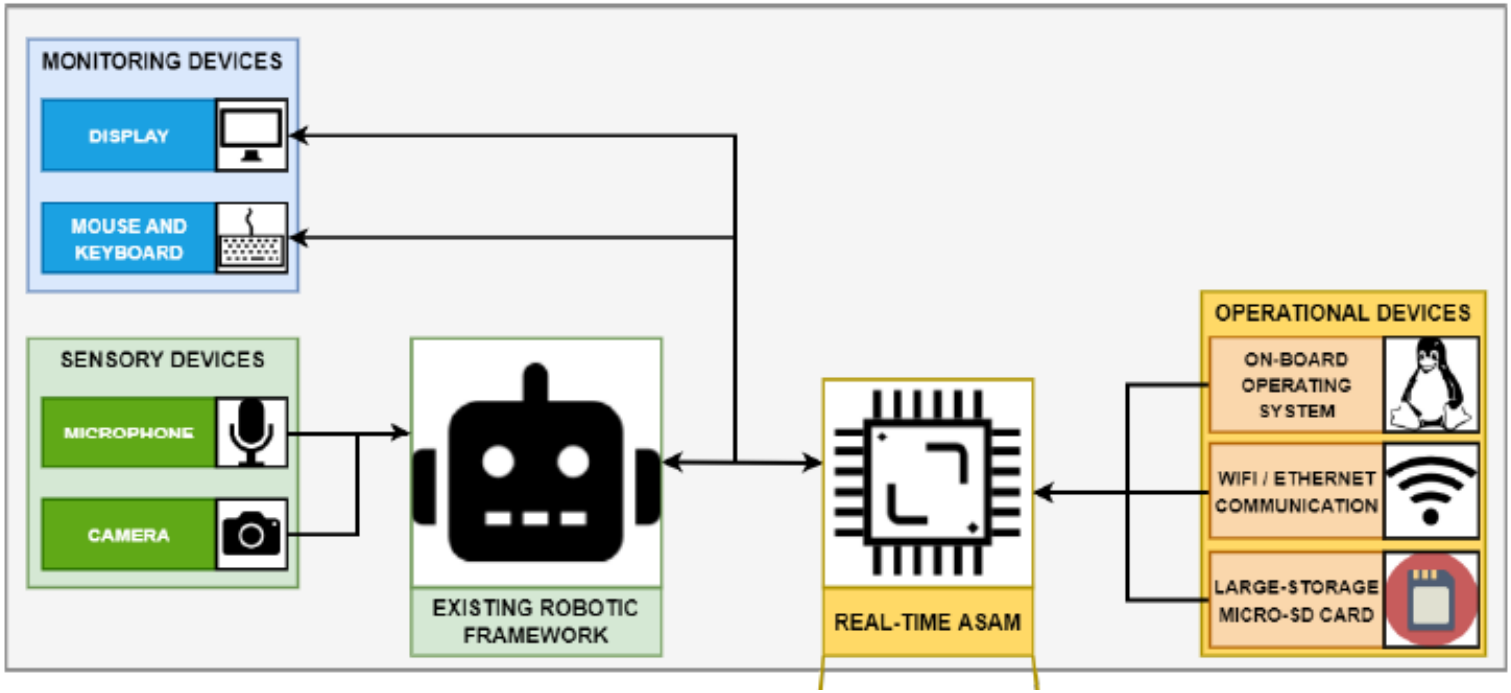
- An incoming sample i is assigned to an existing model M if a match m is found for labelling it as an expression e. For working effectively, models $M_j$, $M_k$, ..., $M_n$ should not be available for matching with the sample i if the model $M_i$ matches the sample i.

MatchModel $\triangleq$ **var** $M$:Model, $i$: sample;
Pre-condition : $M \in$ Model $\wedge$ $i \in$ sample
$\wedge$ IsMatcableTo (IsIncluded$^{-1}(M)$, $i$)
$\wedge$ IsAssignableTo ($M$, $i$)
Post-condition :$\exists e$: Expression $\bullet$ $e \notin$ Expression
$\wedge$ Expression'= Expression $\cup$ $\{e\}$
$\wedge$ Belongs'= Belongs $\cup$ $\{<e, M>\}$
$\wedge$ Matches'= Matches $\cup$ $\{<i, e, >\}$
$\wedge$ IsAssignableTo'= IsAssignableTo – $\cup_{i \in sample}\{< m, i >\}\{e\}$



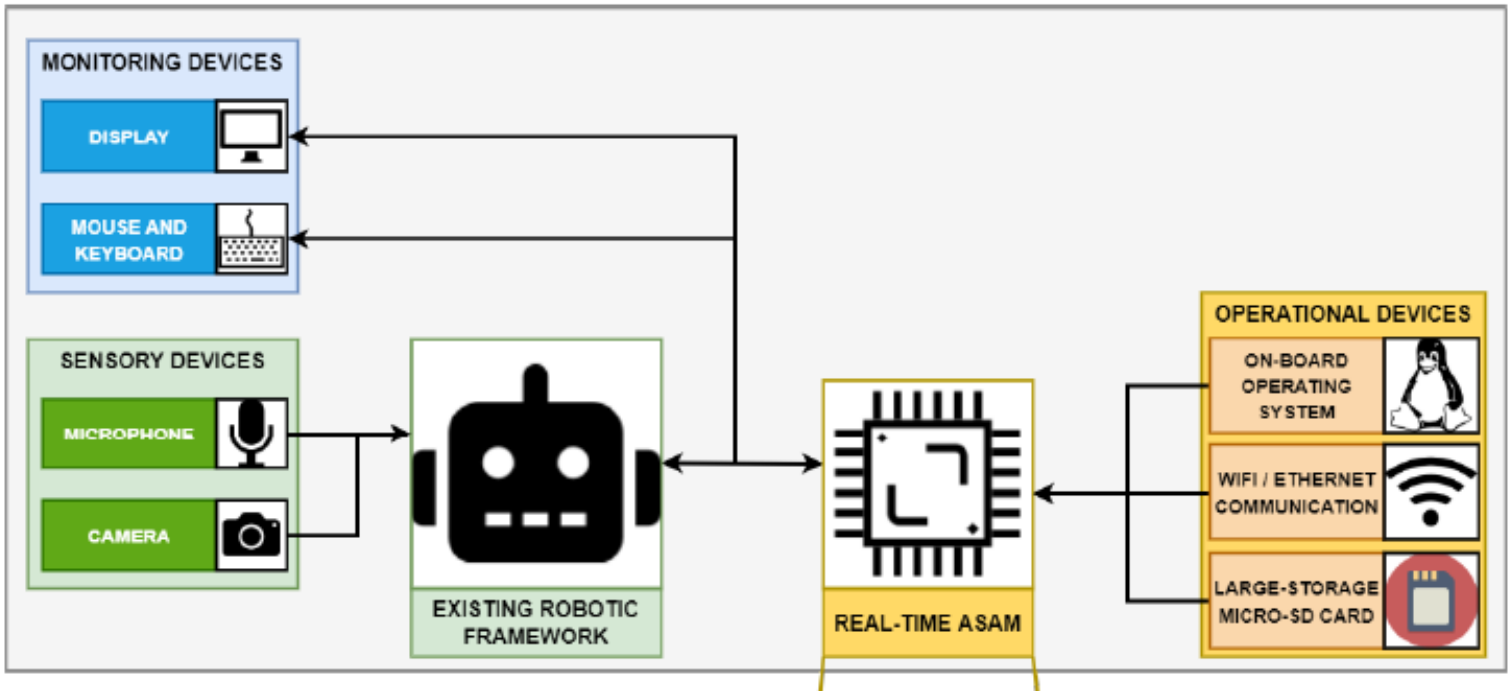
# AXAI capable system design process

- The high-level embedded hardware and functional layout showing its intuitive integration into existing robotic frameworks.
- Operational, sensory, and monitoring devices are integrated to ensure ASAM's full functionality.



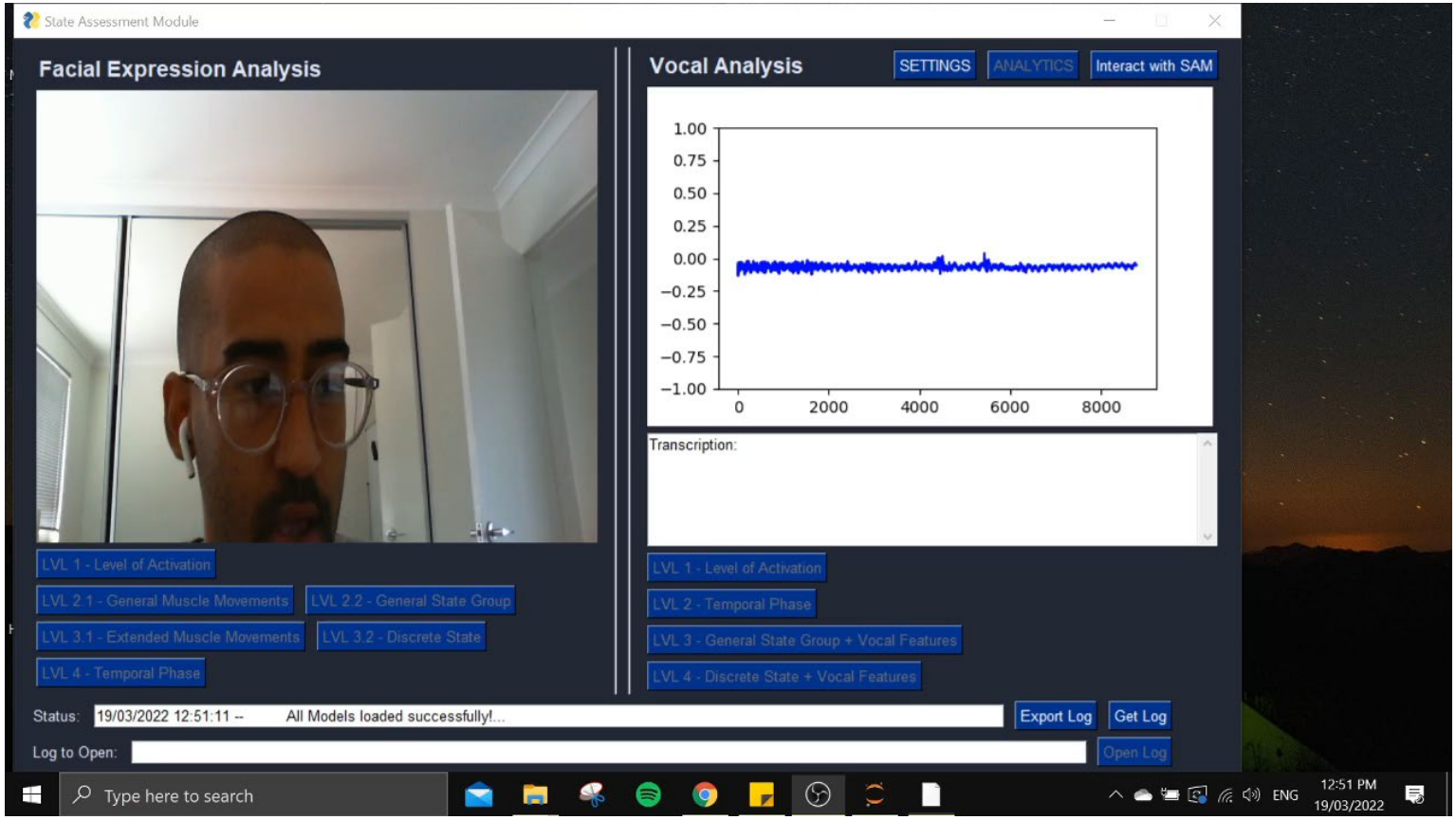
# AXAI capable system design process

- Let’s see the system in operation.



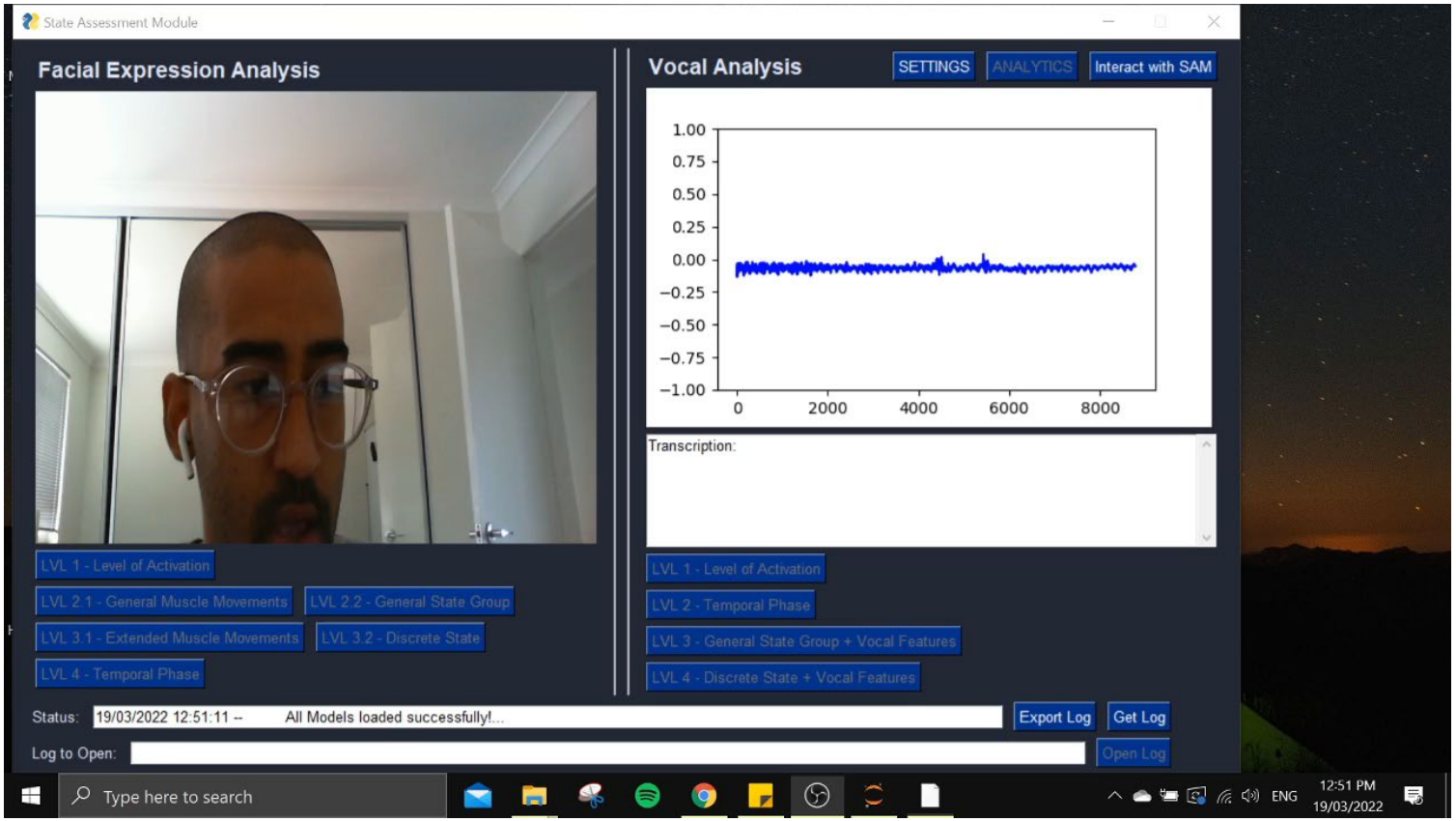
# AXAI capable system design process

- Let’s interact with SAM.



# AXAI capable system design process

- Let’s see another was of interacting with SAM.



# Governance

- We are still not sure what is more problematic, AI in Governance or Governance of AI.

Curtin University

# Questions?