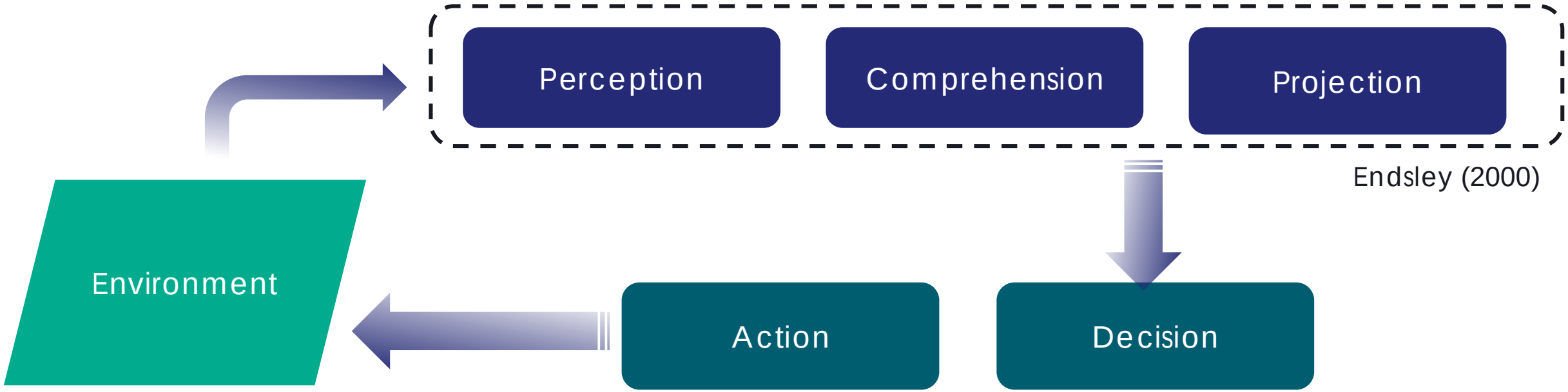
THALES
Building a future we can all trust

Augmenting Autonomous Agents with Expert-Derived Decision Policies

Tanya S.Paul, Stephanie Allogba, Filipe C. Sanches, Jean-Sebastien Thivierge, Daniel Lafond, Antoine Fagette

www.thalesgroup.com

Situational Awareness and Human-Automation Teaming (HAT)



- **Perception: need to process different attributes and dynamic factors of the environment (notice a light or shadow in the sky)**
- **Comprehension: Ability develop an understanding of the situation by making connection between the different clues and events (novice versus expert might interpret an event as more or less threatening)**
- **Projection: Ability to project future outcomes, to be able to make a prediction and estimate how certain event would unfold to help plan the best course of action (neutralize an UAV)**

Team Situational Awareness (TSA)

- Key component of teaming performance
- Attentional narrowing , workload, fatigue, stress, reaction time
- Data overload
- Maladaptive mental model (non-logical reasoning, inappropriate behavior)
- Out-of- the loop & abuse of automation systems

- 33% mishaps caused by humans (Giese & Carr, 2013)

- 67% due to issues with the machine (Giese & Carr, 2013)
- Researchers have demonstrated that mixed initiative target identification where automated agent provides assistance in locating potential targets in visual search space, that performance deteriorates consistently over time (Demir & Likens, 2019)
- They suggest that automated agents who are trained to detect particular stimuli may not perform as well as an alert human (Demir & Likens, 2019)
- Issues with explainability and transparency

SPARQ: Synergic Partners with AI-Reinforced IQ

Perception

> Digital Twin – Perception beyond sensing

- **Situation Awareness**

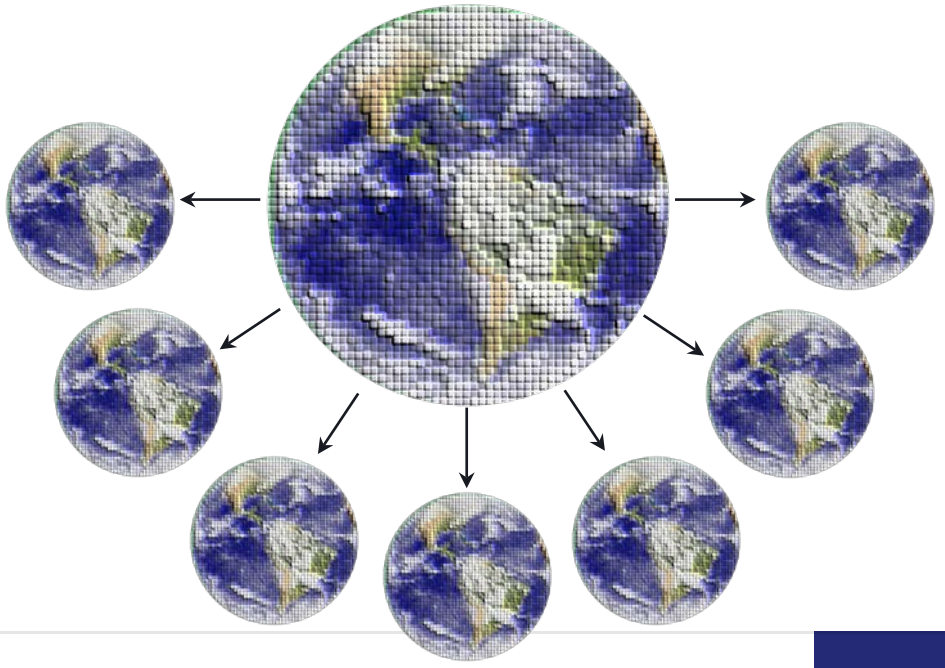
  A virtual representation and understanding of the current situation and its history, based on sensor data, shared data and simulation models

- **Anticipation and What-If Scenarios Exploration**

  A series of forecasts regarding the current situation and how it may unfold depending on various What-If scenarios of interest

- **Enabler for decision support and optimization**

  Leveraging sensor data and simulation models in order to have a complete situation awareness and reliable anticipation insights that can be presented to the operator and processed for decision making



OPEN

Cognitive Shadow

Comprehension

Thales Canada Defense & Security (TCDS) | Thales Land and Air System (LAS) | Air Force Canada & France | Royal Canadian Mounted Police (RCMP)

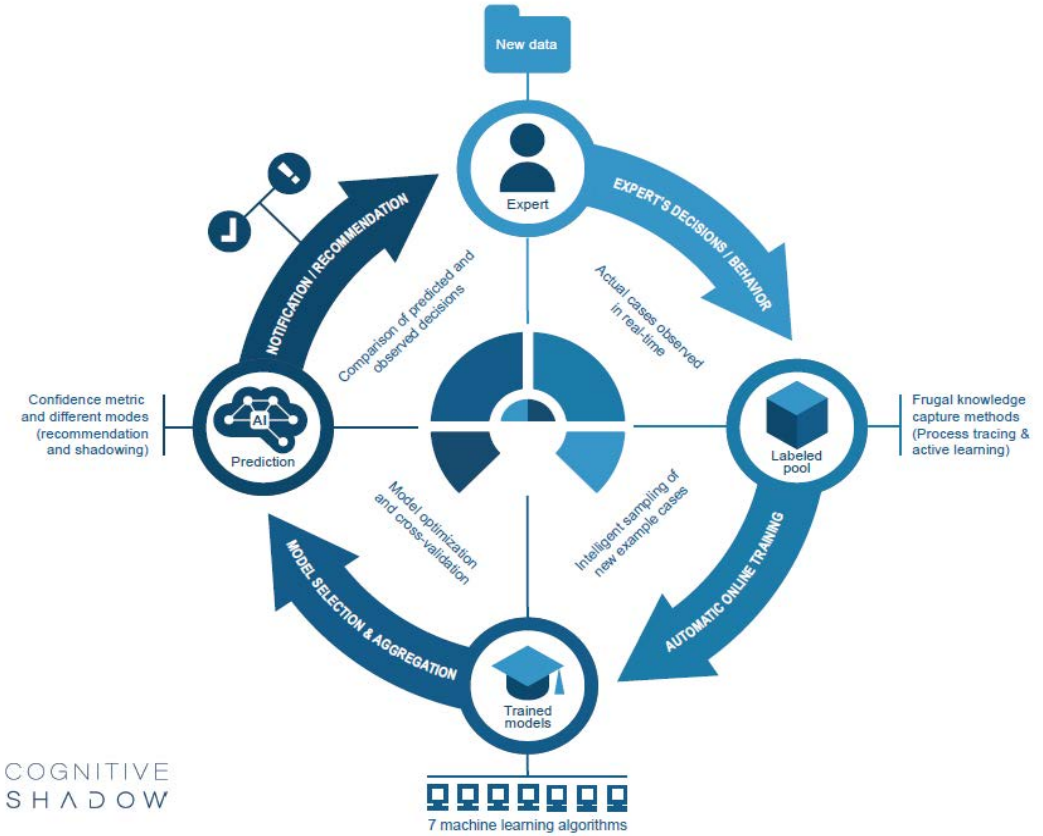
- **Expert-Policy**
  - **Policy capturing**

    Enables interpreting situations by deriving a cognitive model of experts
  - **Real-time recommendation**

    Enable online interaction and learning between agents and humans by accepting or rejecting assessments
  - **Reduced bias and low SA**

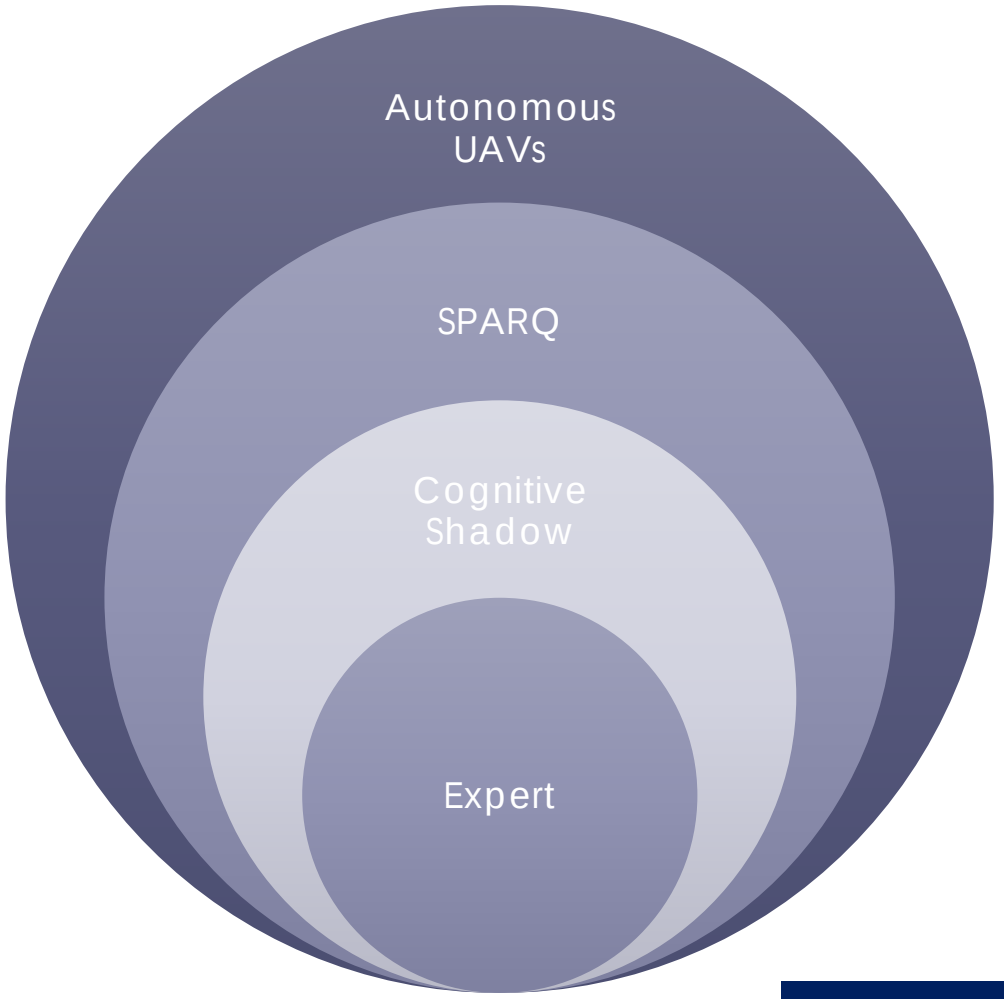
    Enable capturing human expertise in nominal conditions (peak SA without fatigue, stress, overload)
  - **Decision-making**

    Enables human-centric decision-making for the automated UAVs actions



OPEN

Human-Automation Teaming



Autonomous Agents

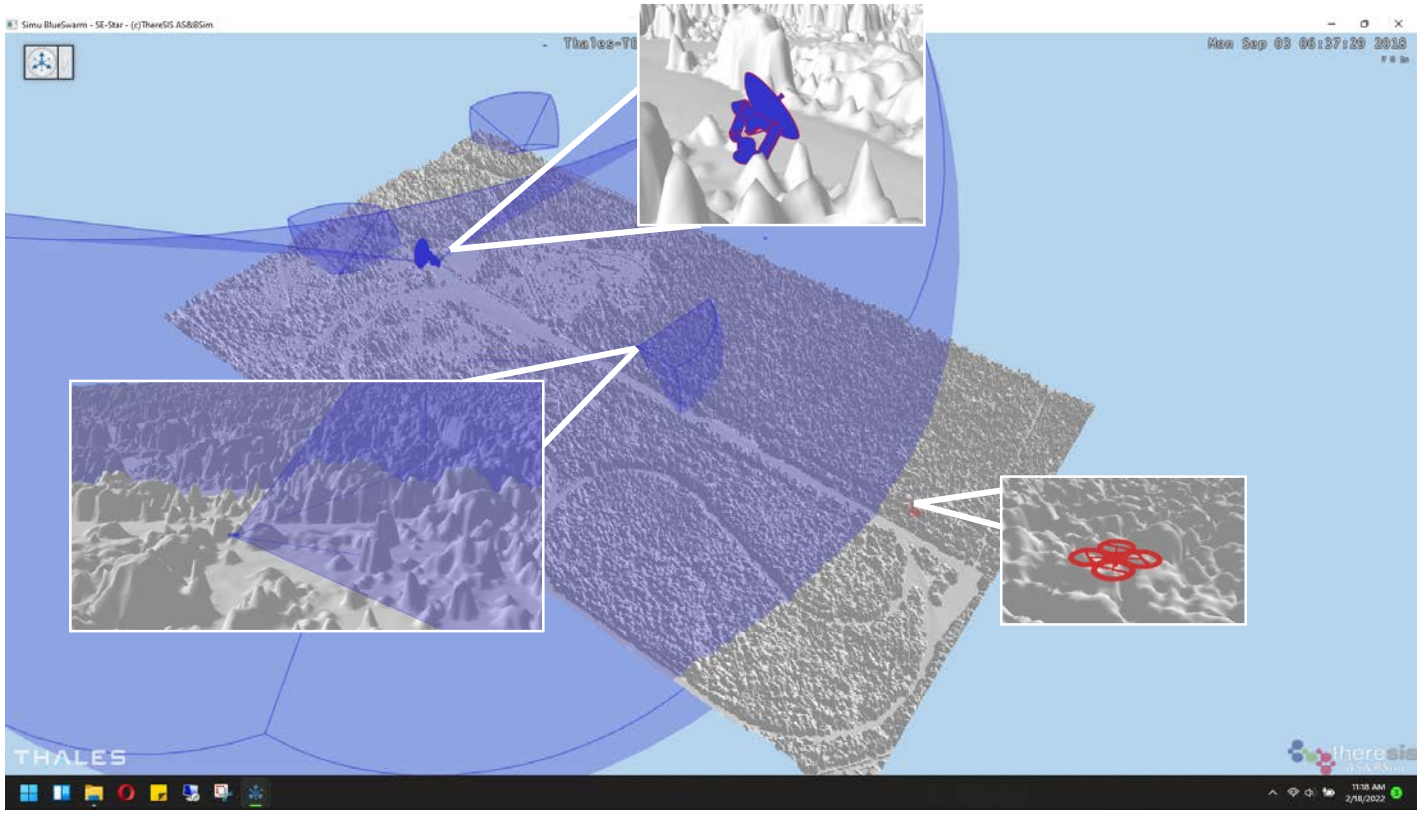
Perception

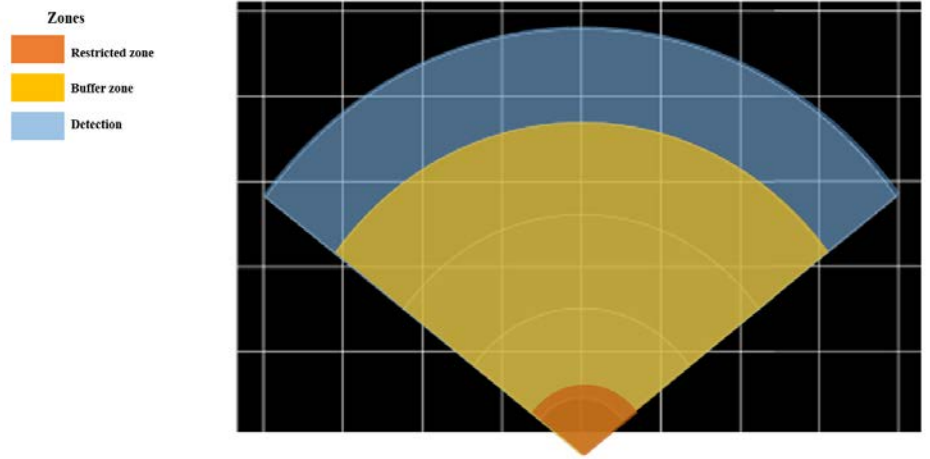
Comprehension

Action and decision-making

Human

Simulated Environment SE-STAR





- Autonomous Robotic Aviation (ARA) + ThereSIS
- Customizable terrain
- Simulated entities (radar, blue force, red force, ground control stations)
- Simulated scenarios
- Testbed for HAT Teaming
- Train, test, deploy and collect data

Threat Level for protected facility

Threat Levels	Definition
Clueless	A UAV is a BLVOS hobbyist testing their drone.
Careless	A UAV considered a supply delivery drone whose flight plan has been defined without consideration of the restricted airspace.
Criminal No Harm	A UAV navigated by a curious photographer getting closer to the restricted area to take pictures.
Criminal with Harm	Is a UAV terrorist-operated drone with an explosive payload.

Expert-Policy for Augmenting AI agent SA capability

Expert

Behavioral Pattern | Detection | Recognition | Identification

SPARQ : Digital Twin

Feature Extraction | Feature Selection | Feature Pre-processing | Feature engineering

Cognitive Shadow

Cognitive Modelling | Knowledge capture | Model Training & Testing | Performance Analysis

Threat Identification Model

Projection

Watch-dog

Stand-by

Neutralization

Actions

SPARQ: Digital Twin

Features	Definition	Feature Engineering	Encoded
Speed	The velocity of the drone measured in m/s.	[Low, Medium, High]	[0,1,2]
Altitude	Position of the UAV with respect to the ground.	[Low, Medium, High]	[0,1,2]
Direction Azimuth	Angular position with respect to the restricted area.	[Right, Left, Front, Rear]	[0,1,2]
Zone	The zone within which the UAV is currently positioned.	[World, Buffer, Restriction]	[0,1,2,3]
Direction Elevation	Angular position with respect to restricted area	[Up, Level, Down]	[0,1,2]
Acceleration	The acceleration of the UAV	[High,Medium, Low Deceleration, No Acceleration, High, Medium, Low Acceleration]	[-3,-2,-1,0,1,2,3]
Distance	The distance of the red UAV from the restricted zone.	[Very Far, Far, Close, Very Close]	[0,1,2,3]
Past Threat Levels	The previous classification based on the expert policy.	[Clueless, Careless, Criminal No Harm, Criminal with Harm]	[0,1,2,3]

Model Training & Testing

Cognitive Shadow

Projection

- **Seven supervised machine learning algorithms**
- **716 synthetic instances to classify**
- **Virtual expert policy used as ground truth**

Decision Tree

Version: 4

Status: READY

Training Time: 8s 821ms

Total Samples Count: 716

Training Window: 100%

Clueless: 100.00%

Careless: 100.00%

CriminalNoHarm: 100.00%

CriminalWithHarm: 100.00%

K-Nearest Neigh...

Version: 4

Status: READY

Training Time: 9s 539ms

Total Samples Count: 716

Training Window: 100%

Clueless: 100.00%

Careless: 99.38%

CriminalNoHarm: 93.55%

CriminalWithHarm: 100.00%

Support Vector ...

Version: 3

Status: READY

Training Time: 2m 17s 849ms

Total Samples Count: 716

Training Window: 100%

Clueless: 100.00%

Careless: 100.00%

CriminalNoHarm: 99.51%

CriminalWithHarm: 99.48%

Logistic Regressi...

Version: 4

Status: READY

Training Time: 5s 136ms

Total Samples Count: 716

Training Window: 100%

Clueless: 100.00%

Careless: 88.89%

CriminalNoHarm: 91.11%

CriminalWithHarm: 79.59%

Naive Bayes

Version: 4

Status: READY

Training Time: 4s 519ms

Total Samples Count: 716

Training Window: 100%

Clueless: 66.81%

Careless: 64.67%

CriminalNoHarm: 71.28%

CriminalWithHarm: 63.13%

Neural network

Version: 3

Status: READY

Training Time: 1m 20s 838ms

Total Samples Count: 716

Training Window: 100%

Careless: 100.00%

Clueless: 100.00%

CriminalNoHarm: 98.00%

CriminalWithHarm: 95.92%

Random Forest

Version: 3

Status: READY

Training Time: 1m 7s 987ms

Total Samples Count: 716

Training Window: 100%

Clueless: 100.00%

Careless: 100.00%

CriminalNoHarm: 95.24%

CriminalWithHarm: 97.85%

OPEN

Discussion & Conclusion

> Key outcome

- Successful proof-of-concept integration

Limitations

- Computer vision reliance (eg., drone versus bird)
- Hacking pre-filtering ("friend" vs "enemy")
- Consensus on expert judgments

Direction for future work

- > Capture broader range of possible behaviour patterns
- > Enhancing simulation complexity
- > Human-in-the-loop testing and co-learning
- > Field testing

✓ Achieve cognitive air superiority with collaborative AI Agent capable of learning from human expertise

THALES
Building a future we can all trust

Thank you for listening

Tanya Sarah Paul

AI Scientist in Applied Cognition