Intelligenza Artificiale per Sistemi Industriali

Lezione 01 — 24 Febbraio 2026

Alice Plebe

www.aliceplebe.com alice.plebe@unitn.it

OBIETTIVI DEL CORSO

- Cos'è l'intelligenza artificiale
- Perché i legami con neuroscienza, filosofia, e sistemi di controllo sono chiave
- Come funziona una rete neurale
- Capire quando e perché usare reti neurali in ambito industriale
- Capire quali tipologie di reti neurali applicare per:
    - identificazione di sistemi
    - segnali e serie temporali
    - controllo e MPC
- Implementare reti in Wolfram Mathematica / PyTorch

PROGRAMMA DEL CORSO (1/3)

- Evoluzione dell'AI
  - Le domande di Turing, fino a oggi
  - Alternanza tra entusiasmo e "AI winters"

- Dalle teorie filosofiche dell'intelligenza, ai paradigmi fondanti dell'AI
  - Due grandi correnti filosofiche (razionalista, empirista)
  - Cinque definizioni di AI (simbolica, connessionista, Bayesiana, Darwiniana, apprendimento per rinforzo)

- Neuroscienza come ispirazione per le reti artificiali
  - Dal neurone biologico, sinapsi, potenziale d'azione alle ANNs
  - Hebbian learning

PROGRAMMA DEL CORSO (2/3)

- Evoluzione da reti "shallow" a reti "deep"
  - Perceptron, Neocognitron, Boltzmann Machine, Self-Organizing Maps
  - Parallel Distributed Processing, Backpropagation, Stochastic Gradient Descent
  - Rinascita con deep learning
  - Effetto black-box e explainability

- Architetture moderne
  - Paradigmi di training, supervised e unsupervised
  - Reti ricorrenti (Elman, Jordan, LSTM, GRU) e sistemi dinamici
  - Reti convolutive (1D e 2D) per segnali e immagini
  - Transformer (attention mechanism) e serie temporali
  - Reinforcement learning (Q-learning, actor–critic, PPO)
  - Spiking neural networks

PROGRAMMA DEL CORSO (3/3)

- Applicazioni industriali
  - Reti neurali e identificazione di sistemi
  - Reti neurali e Kalman filtering
  - Reti neurali per Model Predictive Control
  - … and more

- Framework di implementazione
  - Python puro (numpy)
  - PyTorch (o Keras)
  - Wolfram Mathematica

MATERIALE

www.aliceplebe.com/teach

+ MOODLE

MODALITÀ D'ESAME (DA CONFERMARE)

- **Prova scritta:** domande a risposta multipla
  - teoria e concetti fondamentali
  - comprensione delle architetture di reti neurali
  - interpretazione di frammenti di codice

- **Progetto implementativo (opzionale):** migliora il voto finale
  - Sviluppo di un modello neurale
  - Python oppure Mathematica

1 Go to wooclap.com

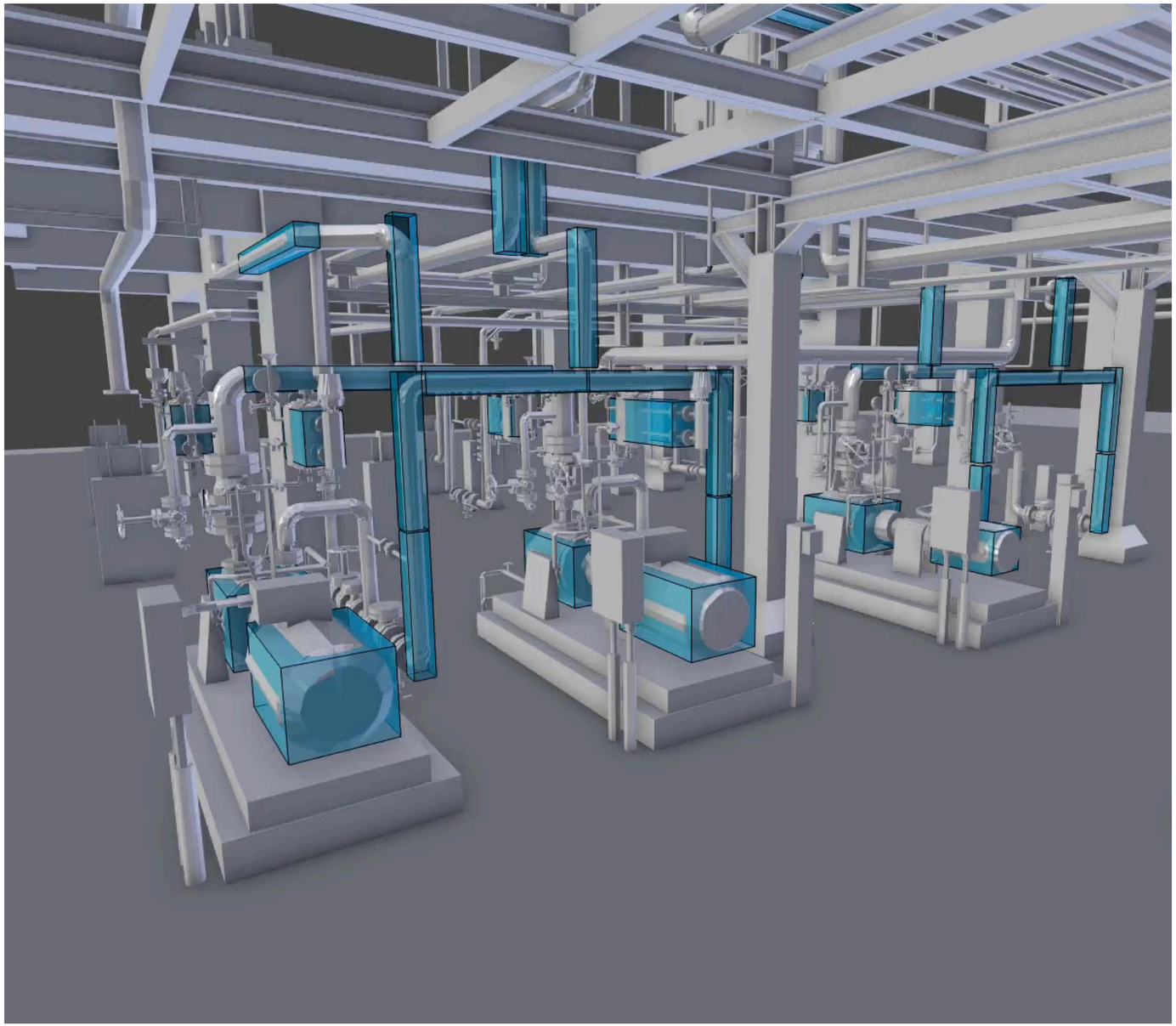
2 Enter the event code in the top banner

Event code

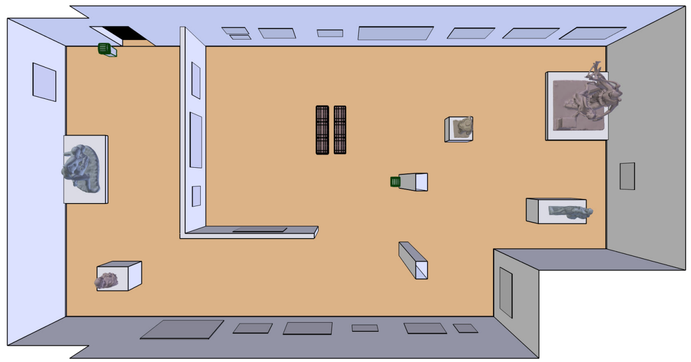
ODBYJZ

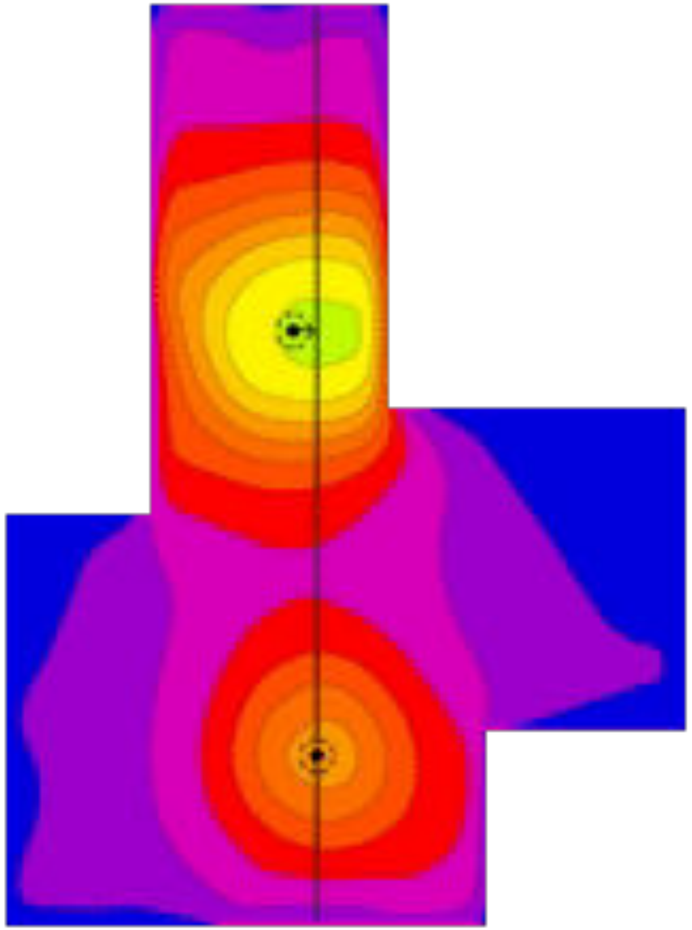
Background

SIMULATING FIRE OUTBREAKS IN INDUSTRIAL PLANTS

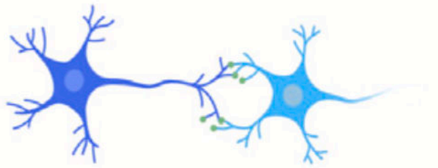


GENETIC ALGORITHM FOR INTERIOR LIGHTING DESIGN

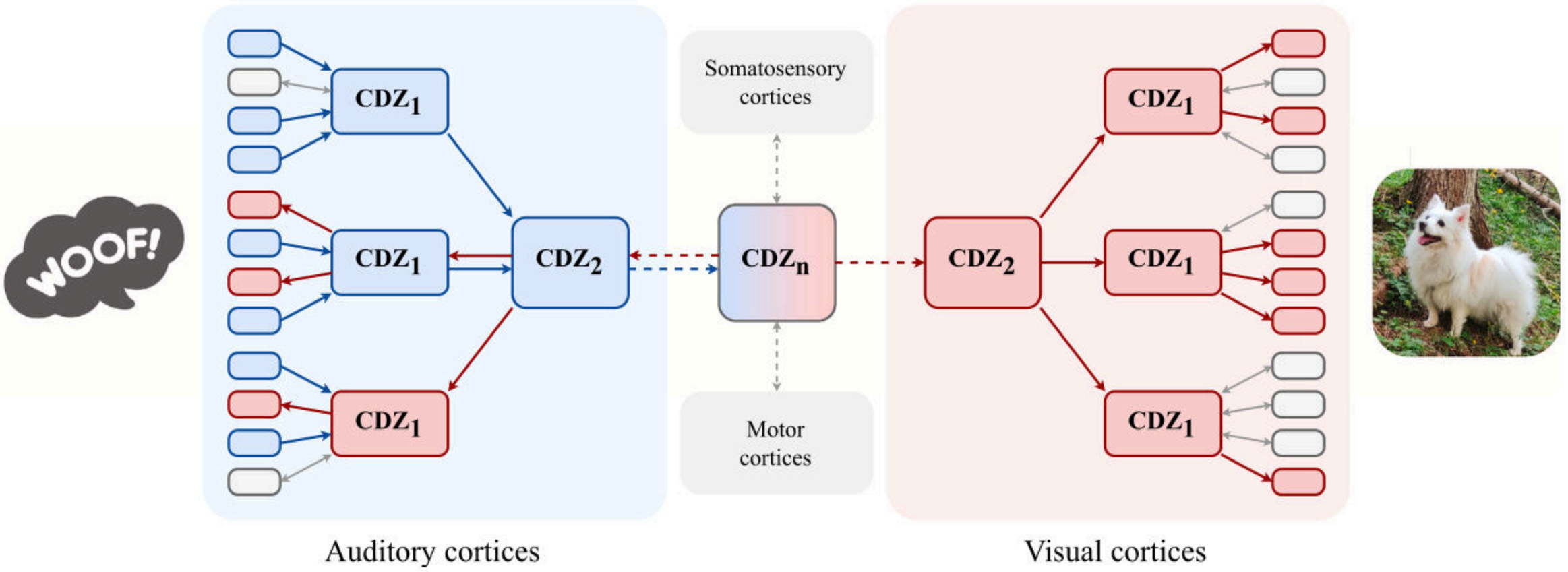




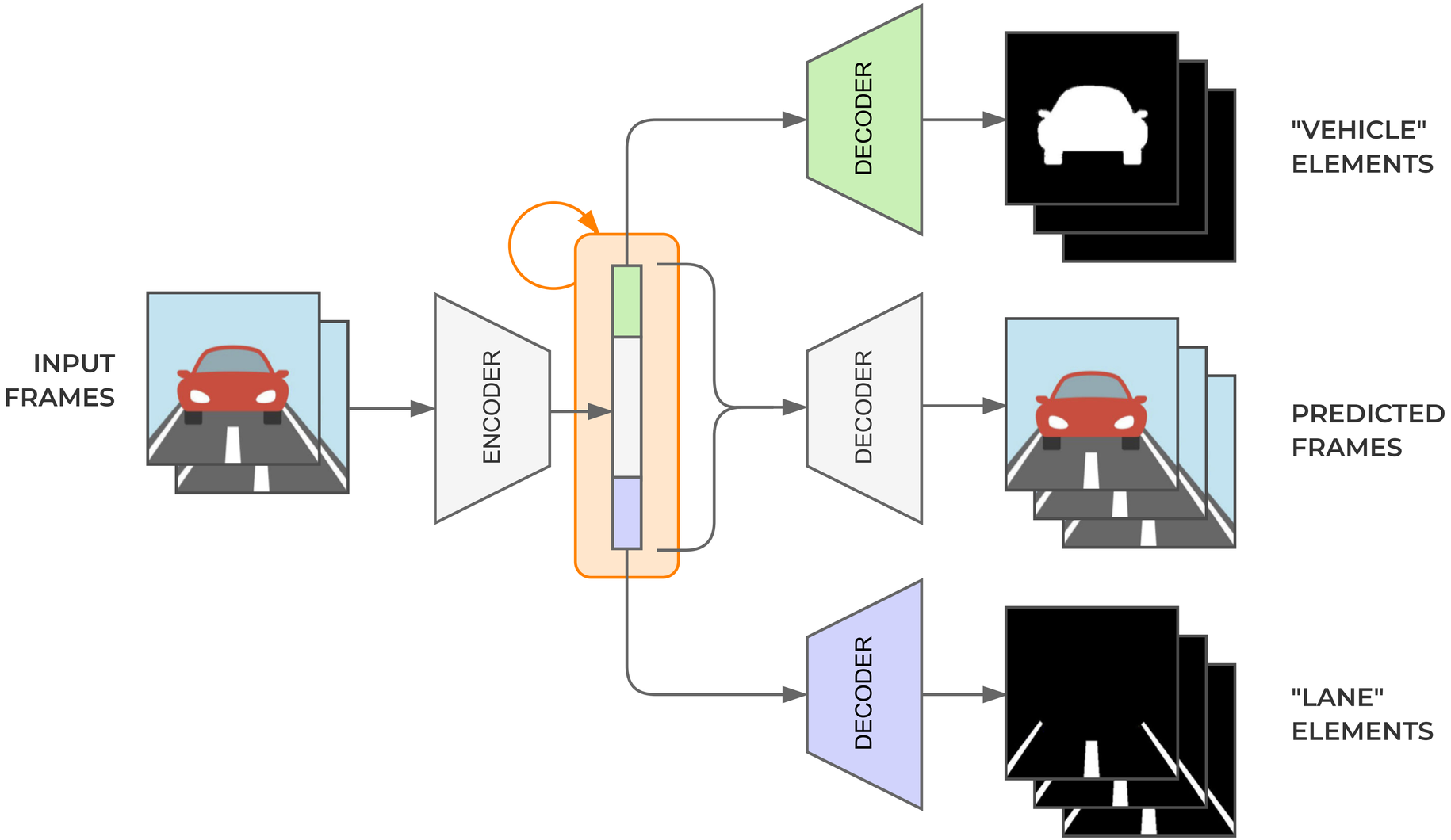
HUMAN-INSPIRED AUTONOMOUS DRIVING

 **(A) BEHAVIORAL INSPIRATION**	 **(B) FUNCTIONAL INSPIRATION**	 **(C) ARCHITECTURAL INSPIRATION**	 **(D) CELLULAR INSPIRATION**
Behavioral cloning, imitation learning {17}	Human-like system parametrization {2}	Visual cortex {2}	Retinomorphic cameras {5}
Human-like constraints of decision space {2}	Attention, social collaboration, emotions {14}	Cerebellum, serotonin {7}	Place cells, grid cells {2}
Observation of external agents {2}	ACT-R, consciousness, microdecisions {7}	Basal ganglia {4}	Spiking neural networks {4}

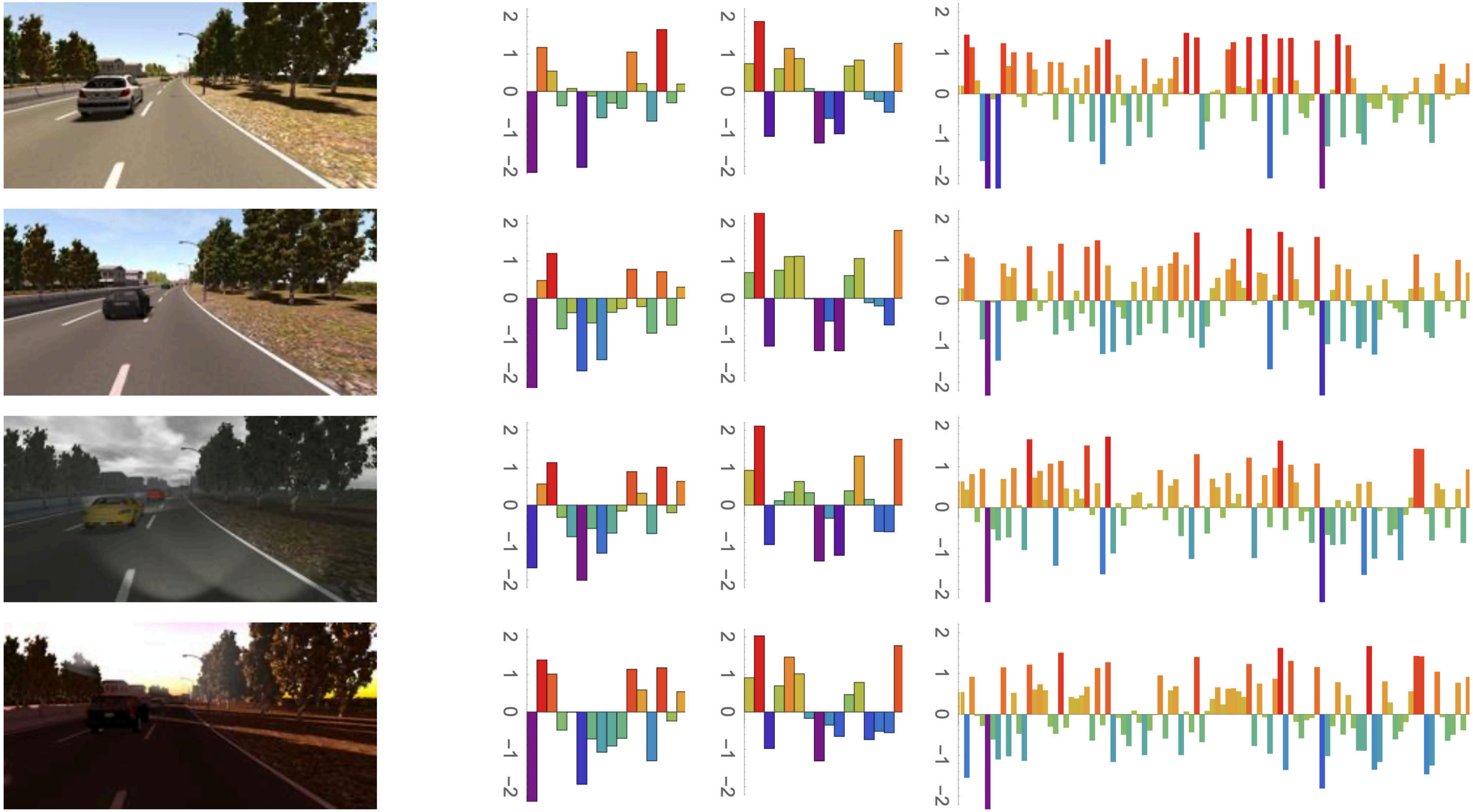
CONVERGENCE-DIVERGENCE ZONES



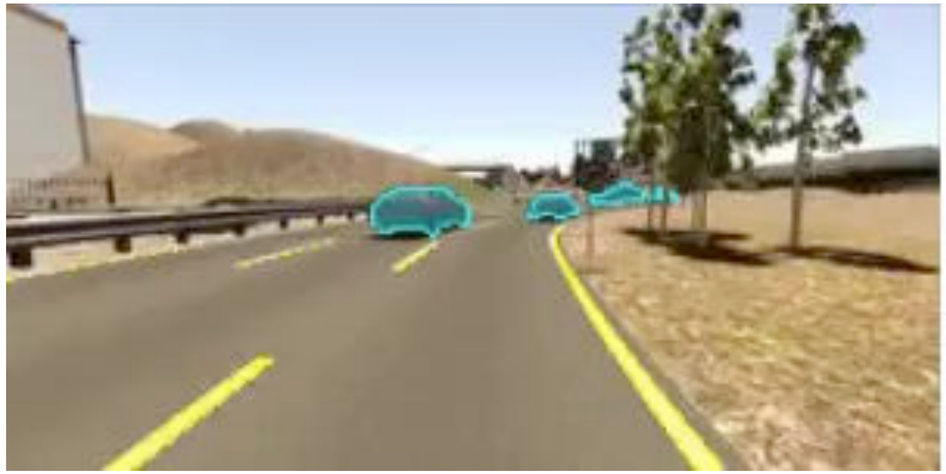
DISENTANGLED LATENT REPRESENTATIONS

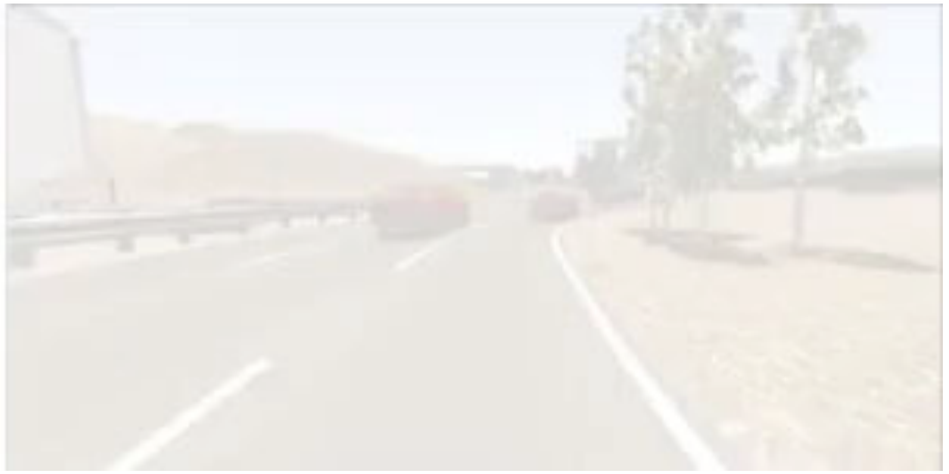


INPUT FRAMES	VEHICLES	LANES	OTHER VISUAL FEATURES

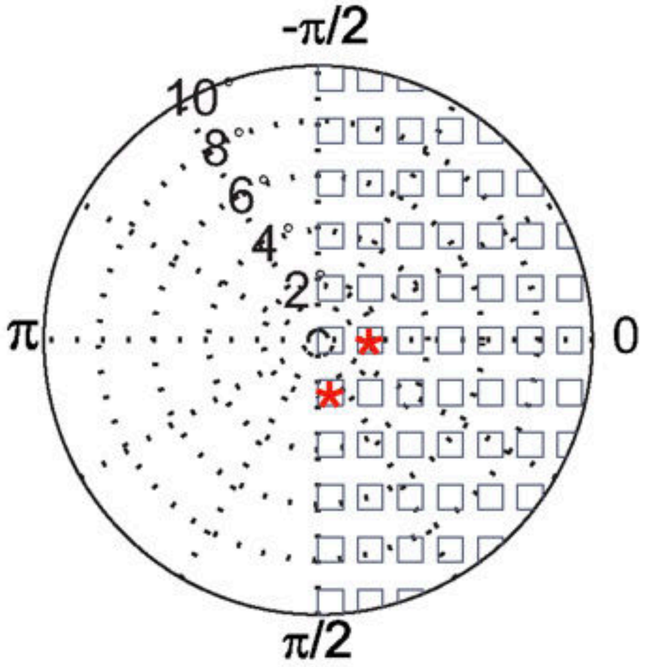


ARTIFICIAL MENTAL IMAGERY

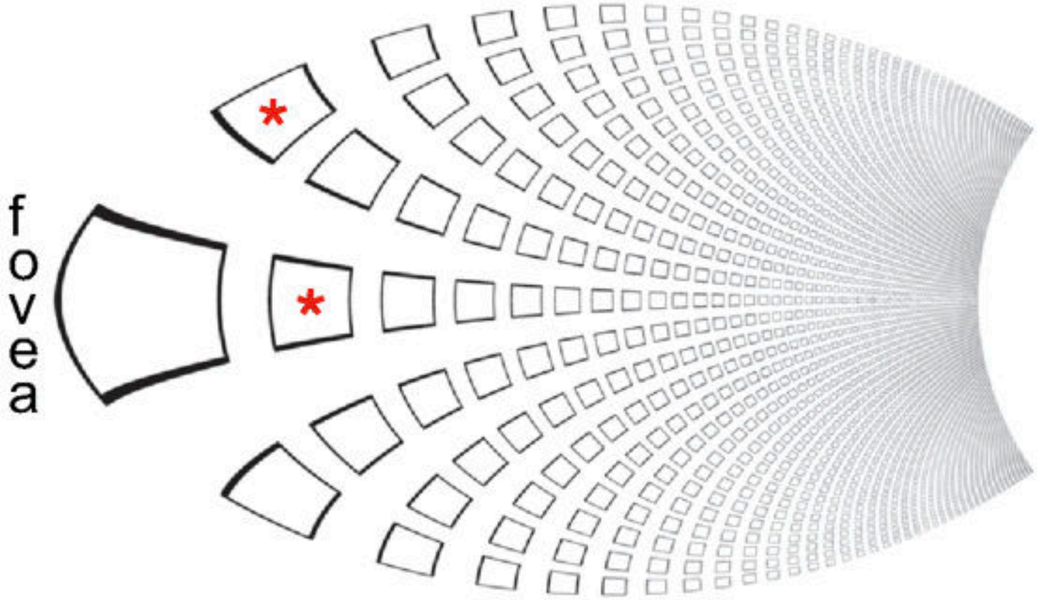




OCCUPANCY GRID MAPPING WITH CORTICAL MAGNIFICATION

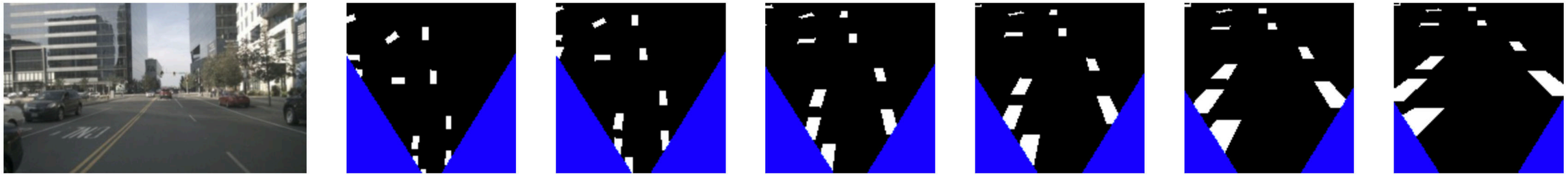


Visual Field



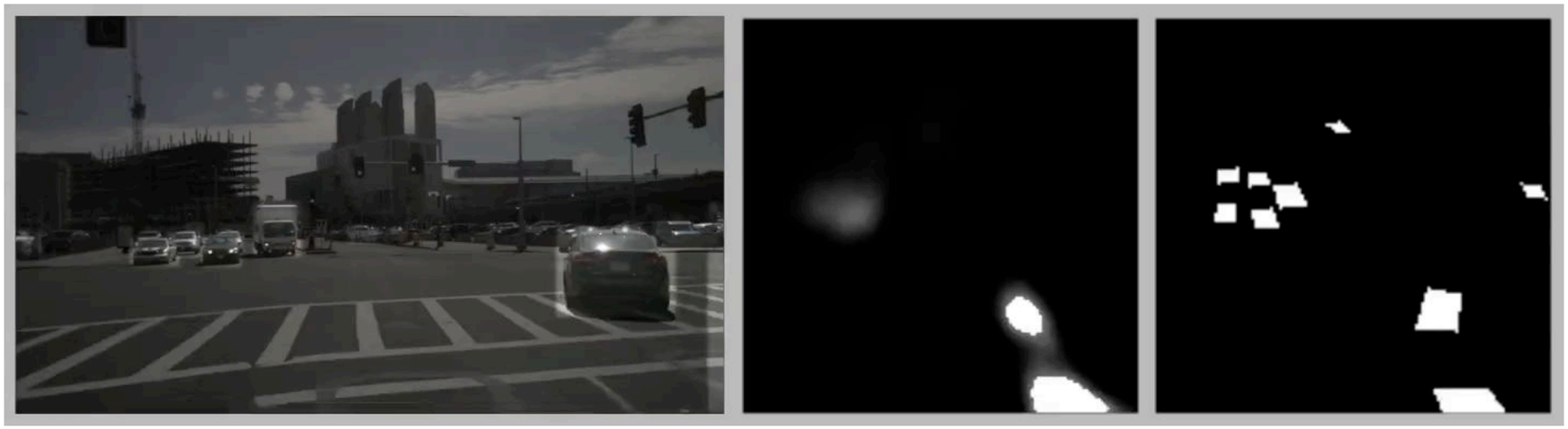
Left Visual Cortex (V1)

OCCUPANCY GRID MAPPING WITH CORTICAL MAGNIFICATION



ORIGINAL

INCREASING FACTOR OF MAGNIFICATION →

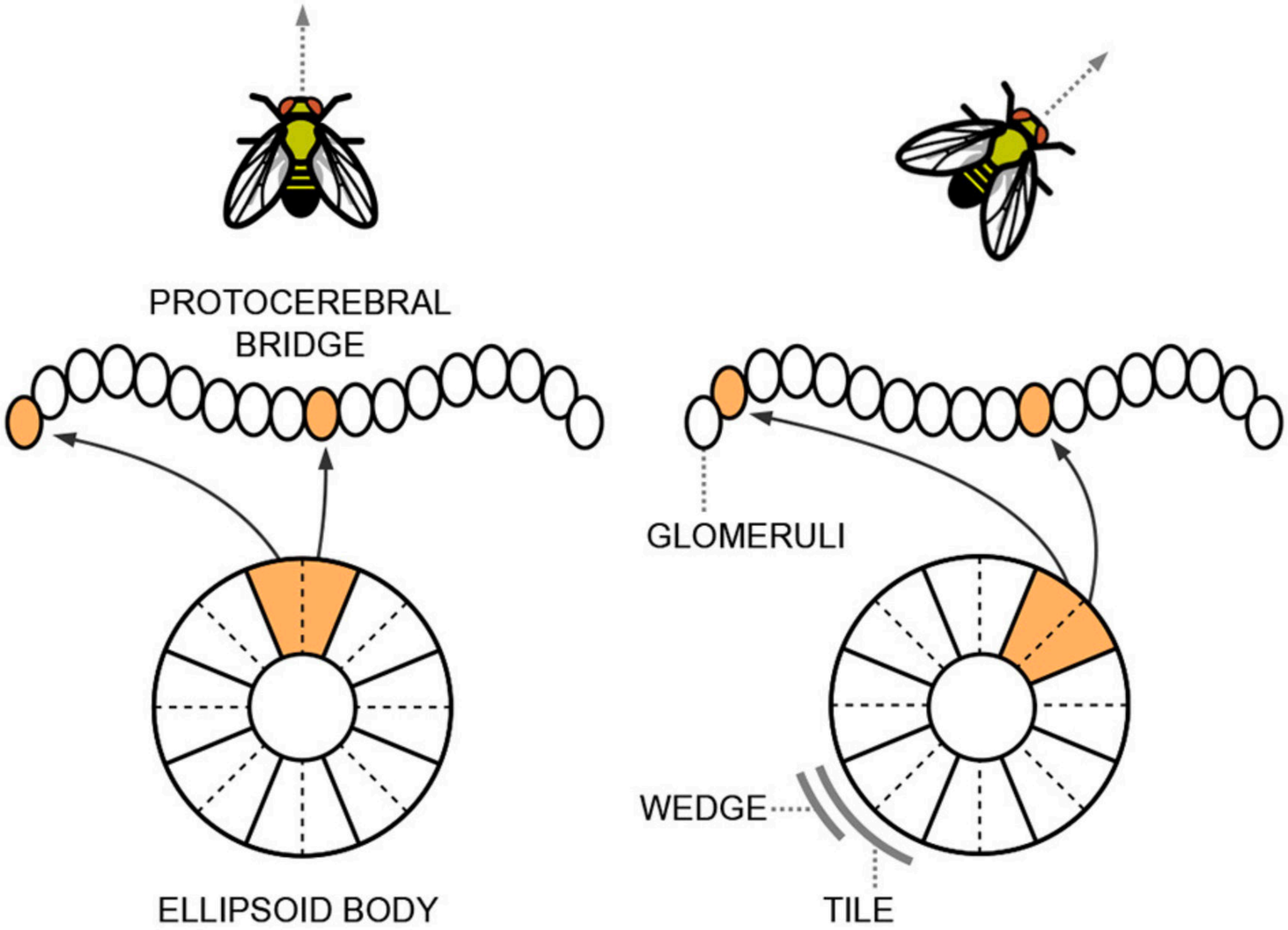


INPUT SEQUENCE
(MONOCAMERA)

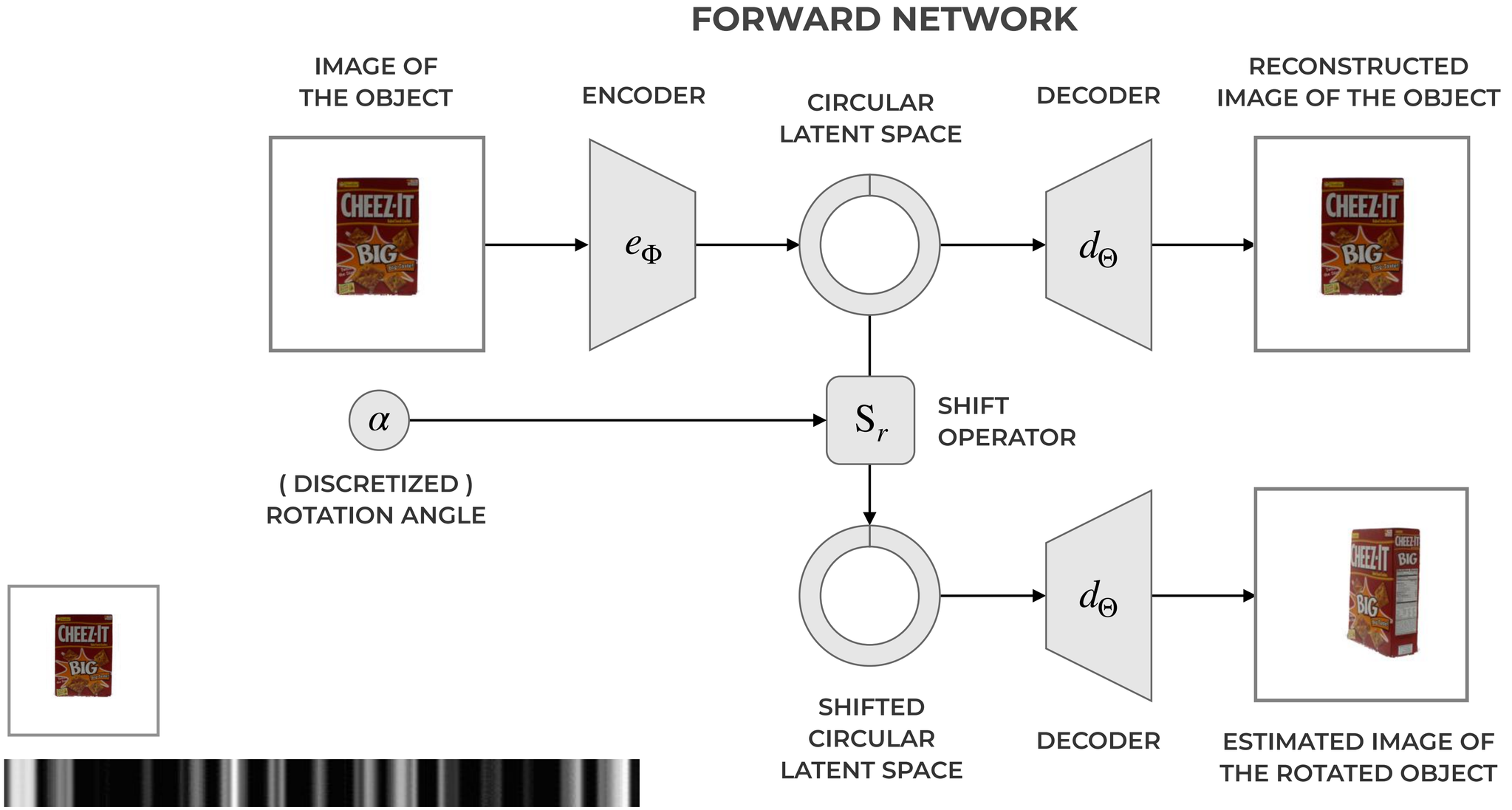
PREDICTED SEQUENCE

TARGET SEQUENCE

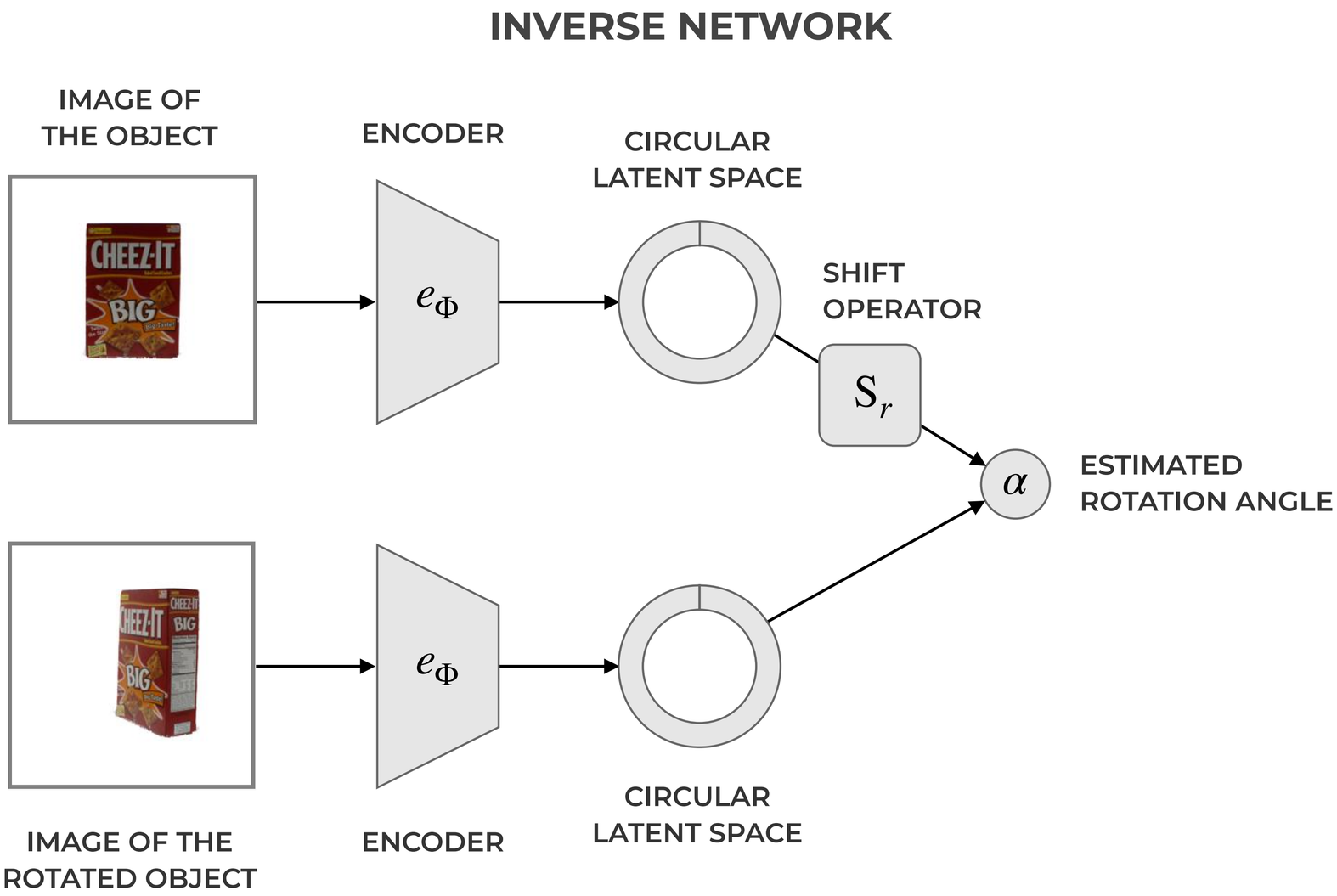
EMBEDDING ROTATION IN CIRCULAR LATENT REPRESENTATION



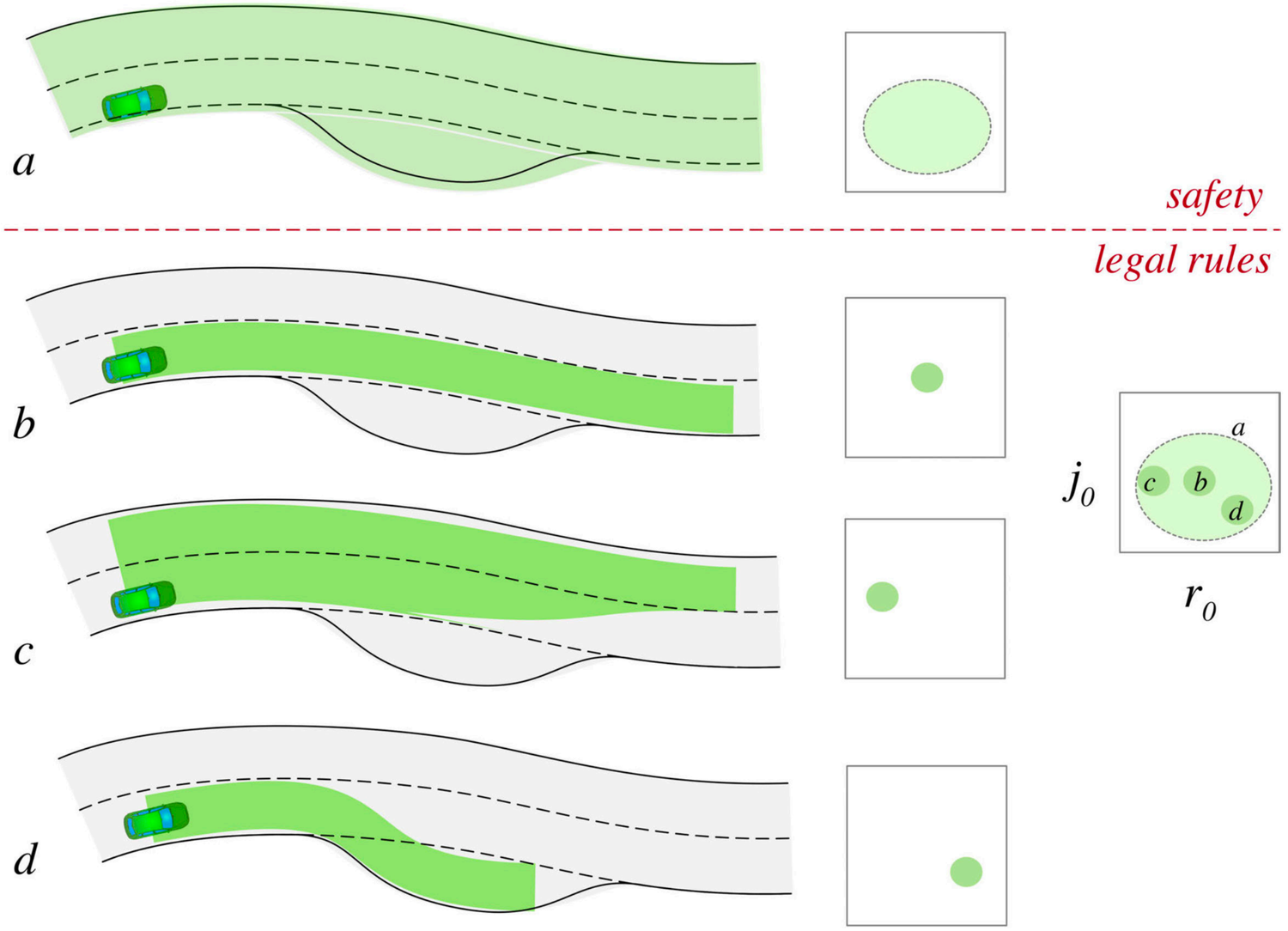
EMBEDDING ROTATION IN CIRCULAR LATENT REPRESENTATION



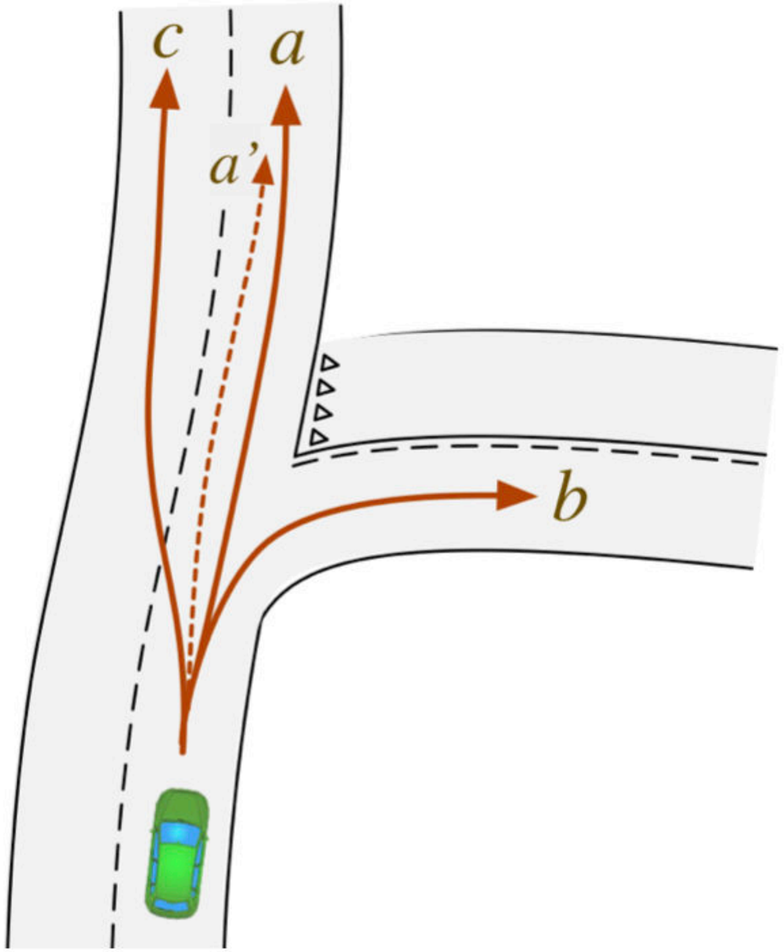
EMBEDDING ROTATION IN CIRCULAR LATENT REPRESENTATION

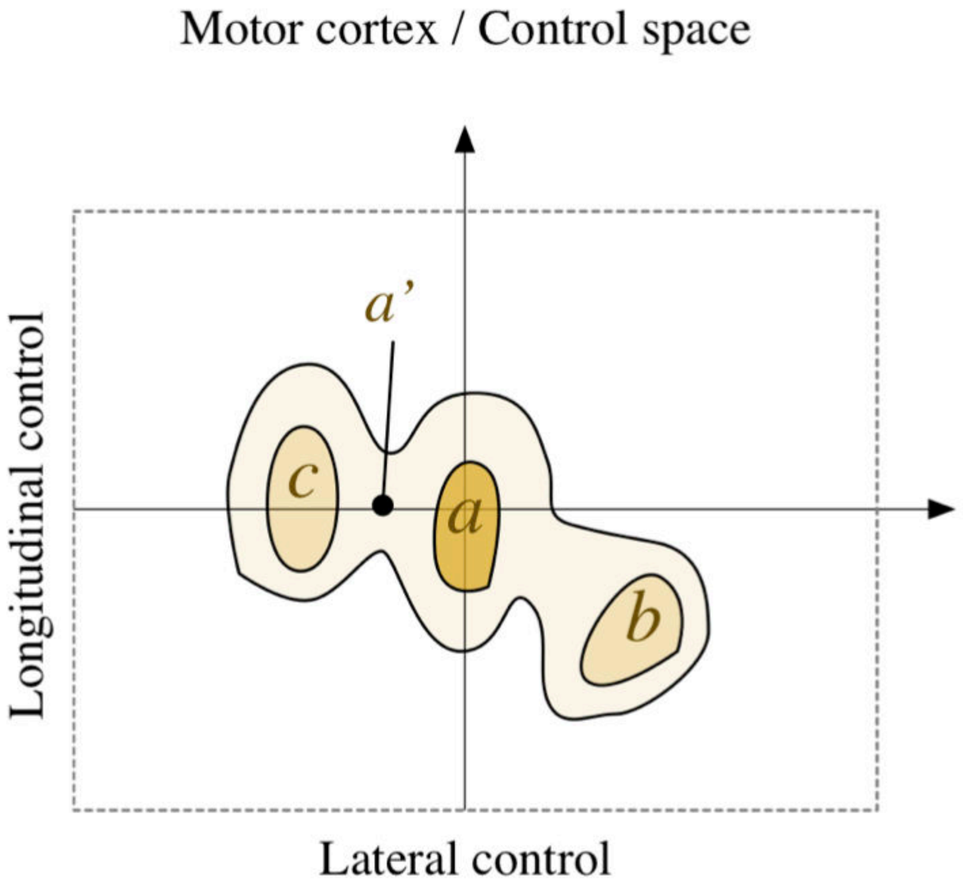


DRIVING ACTIONS AS AFFORDANCES

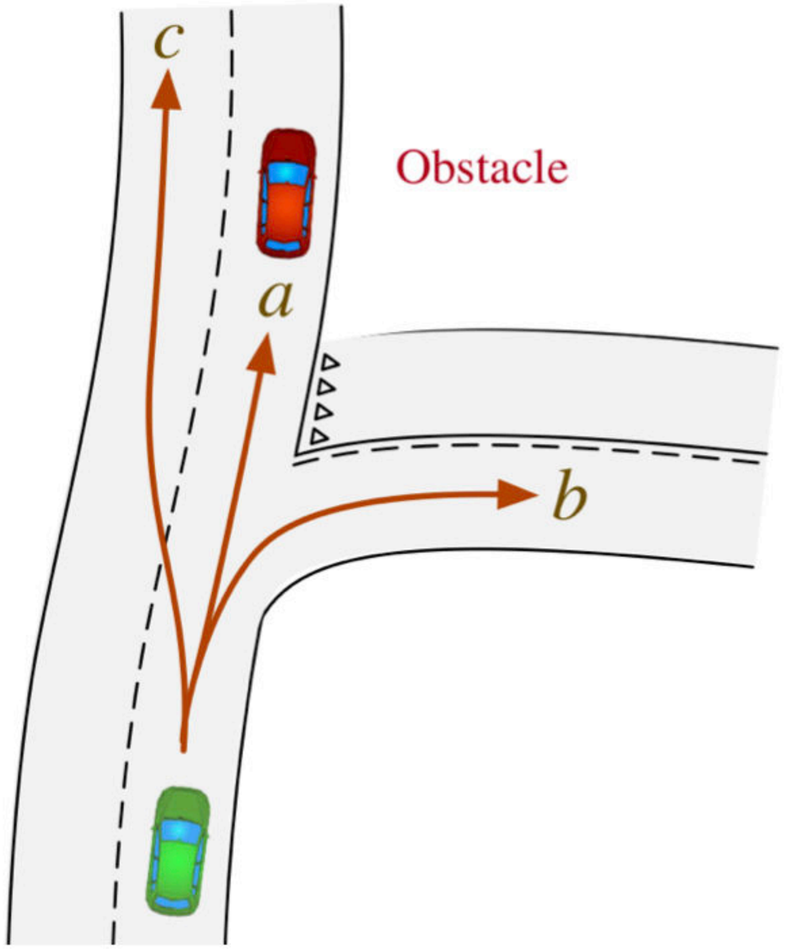


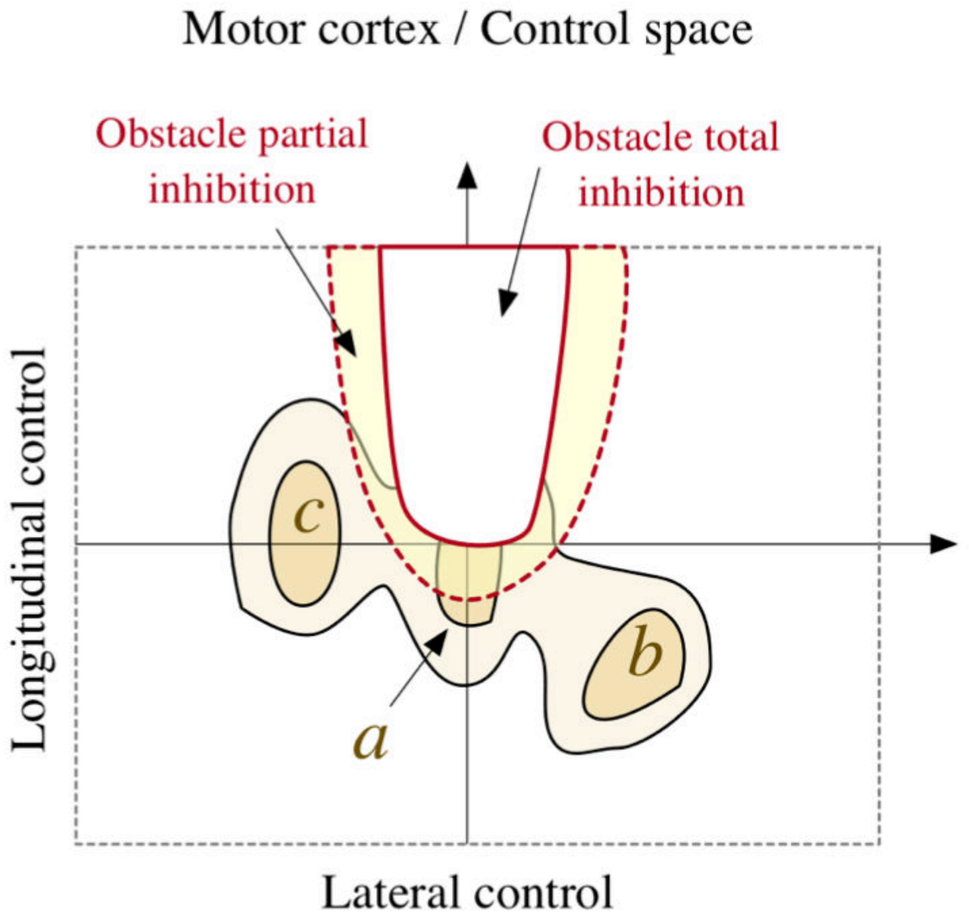
DRIVING ACTIONS AS AFFORDANCES



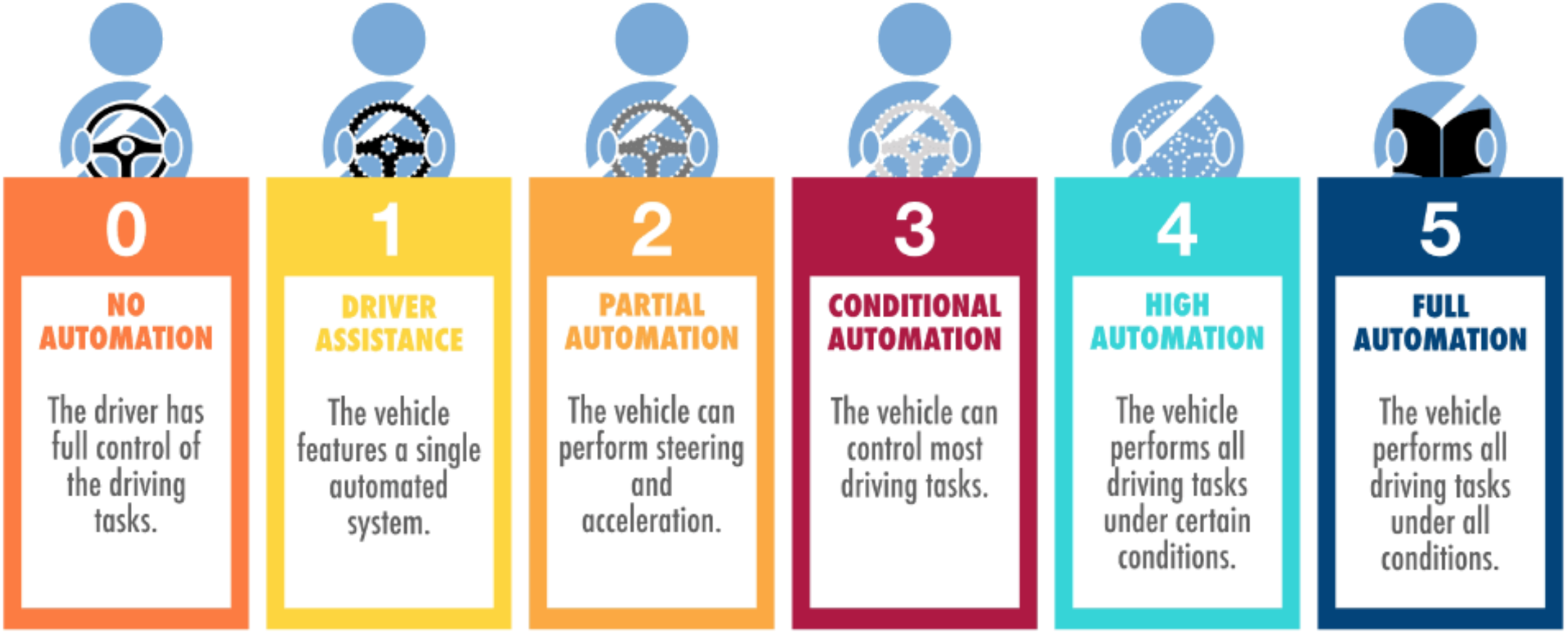


DRIVING ACTIONS AS AFFORDANCES



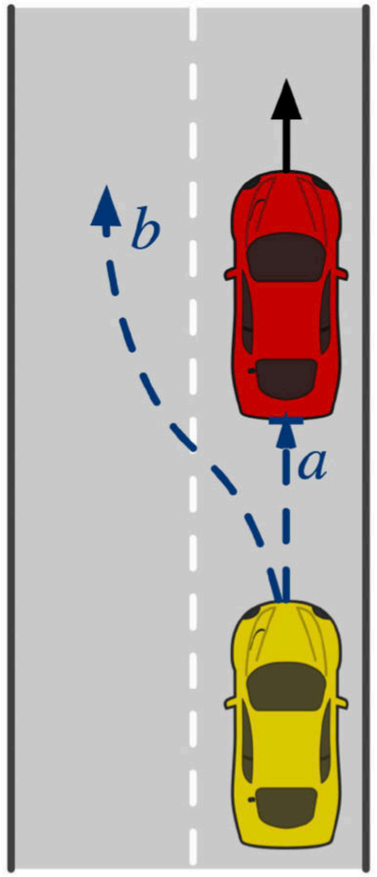


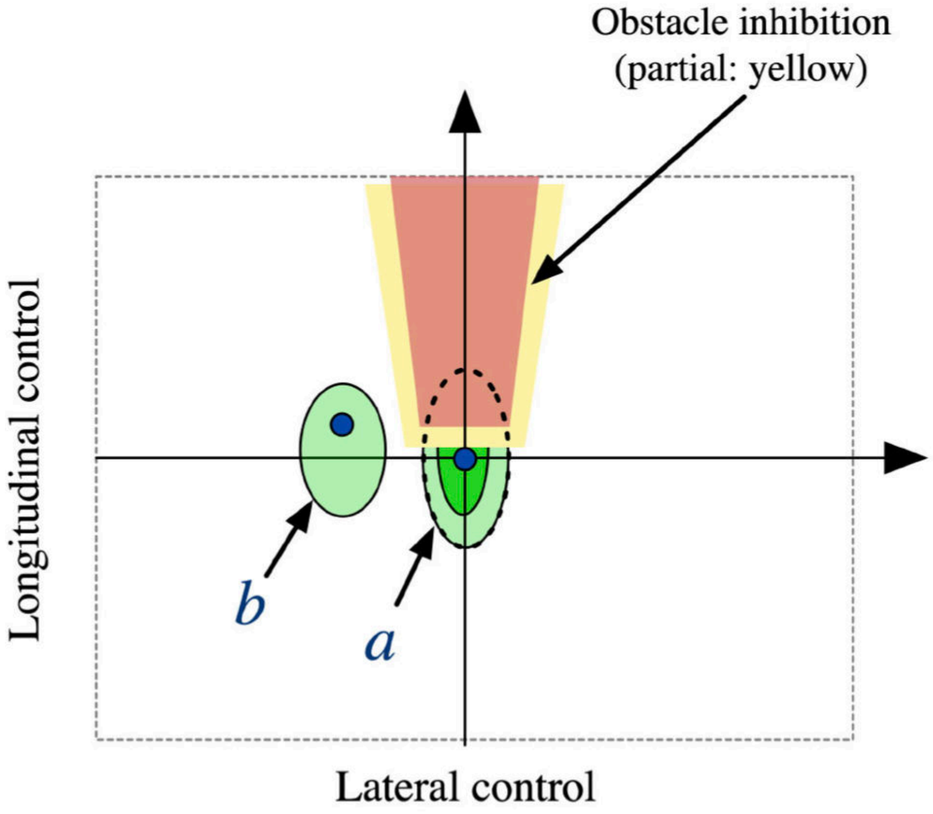
SAE LEVELS OF AUTONOMOUS DRIVING



DISTRIBUTED COGNITION

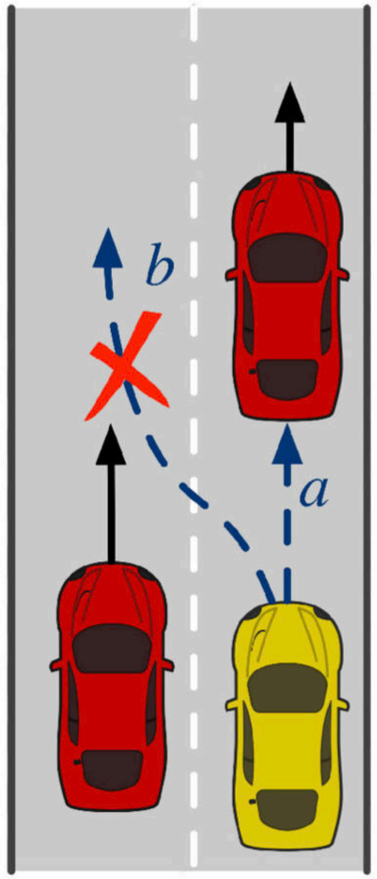
HORSE-RIDER METAPHOR

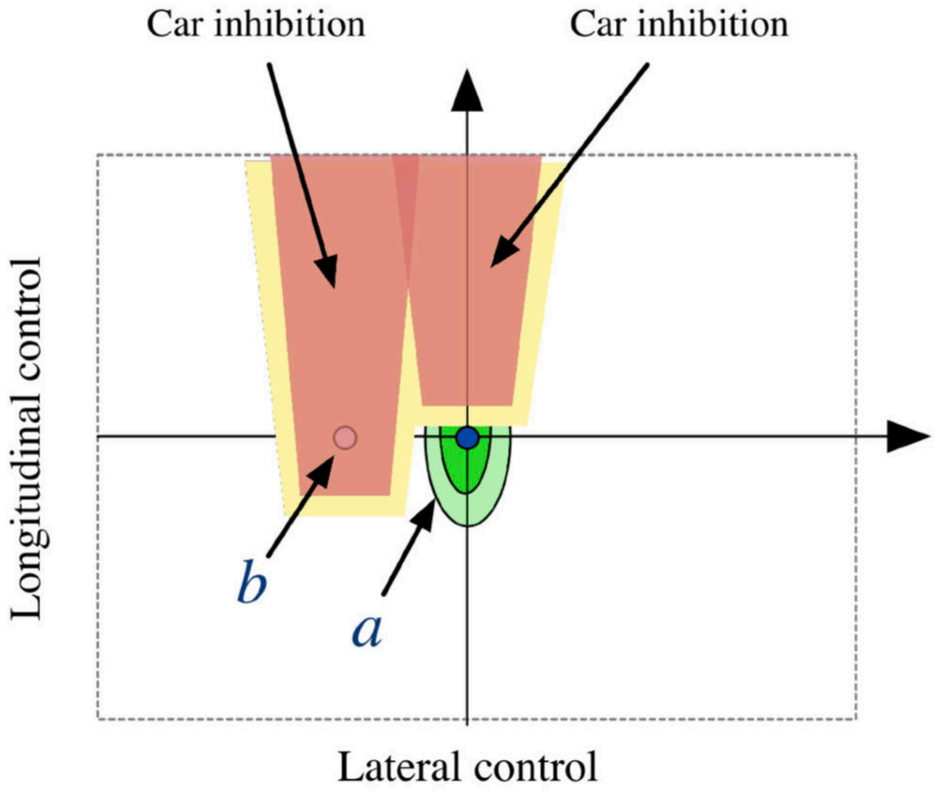




DISTRIBUTED COGNITION

HORSE-RIDER METAPHOR

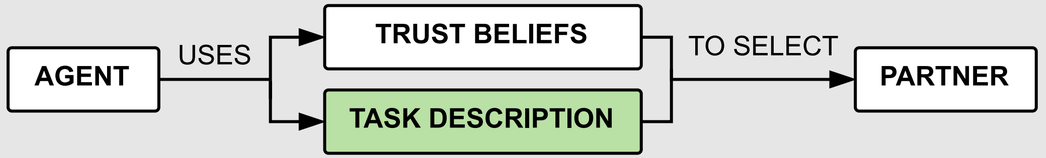




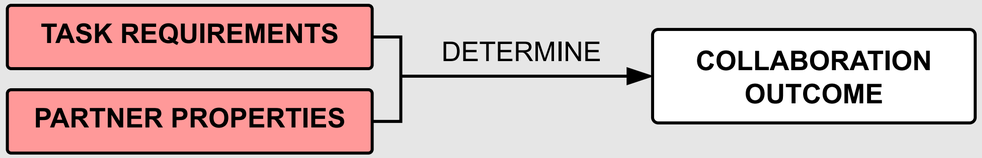
TRUST AS COLLABORATION MECHANISM IN LANGUAGE MODELS

1.

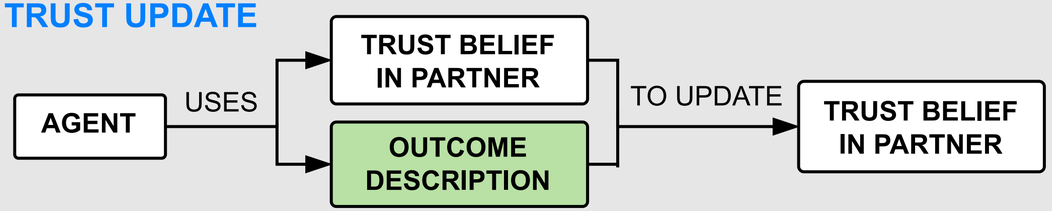
2. PARTNER SELECTION



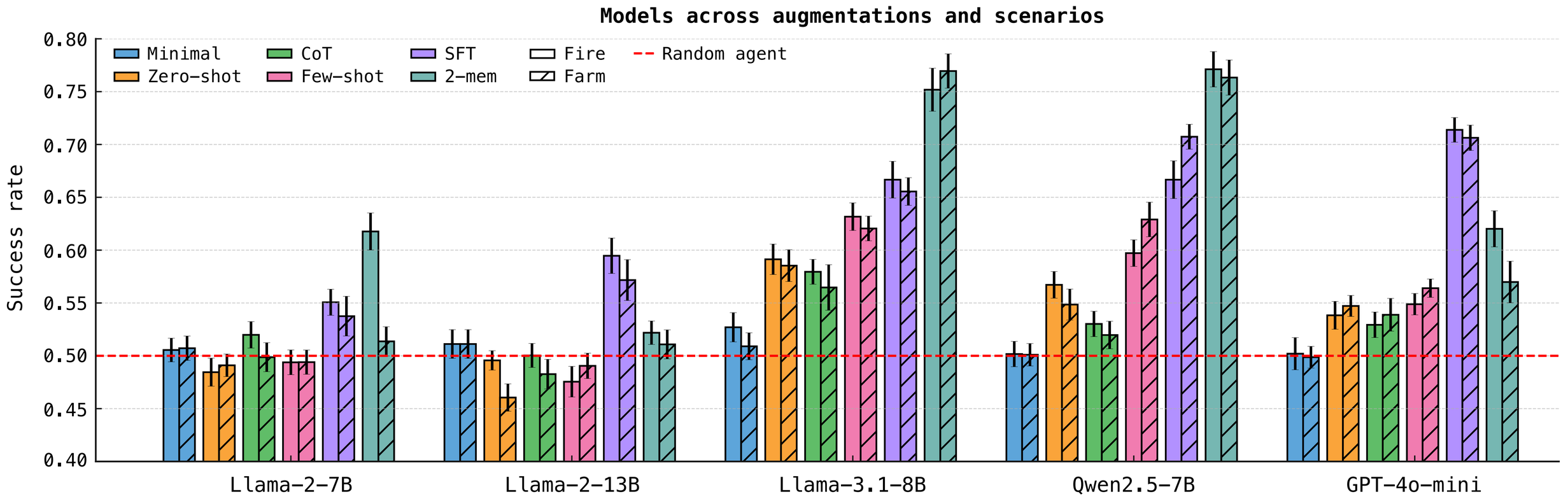
3.



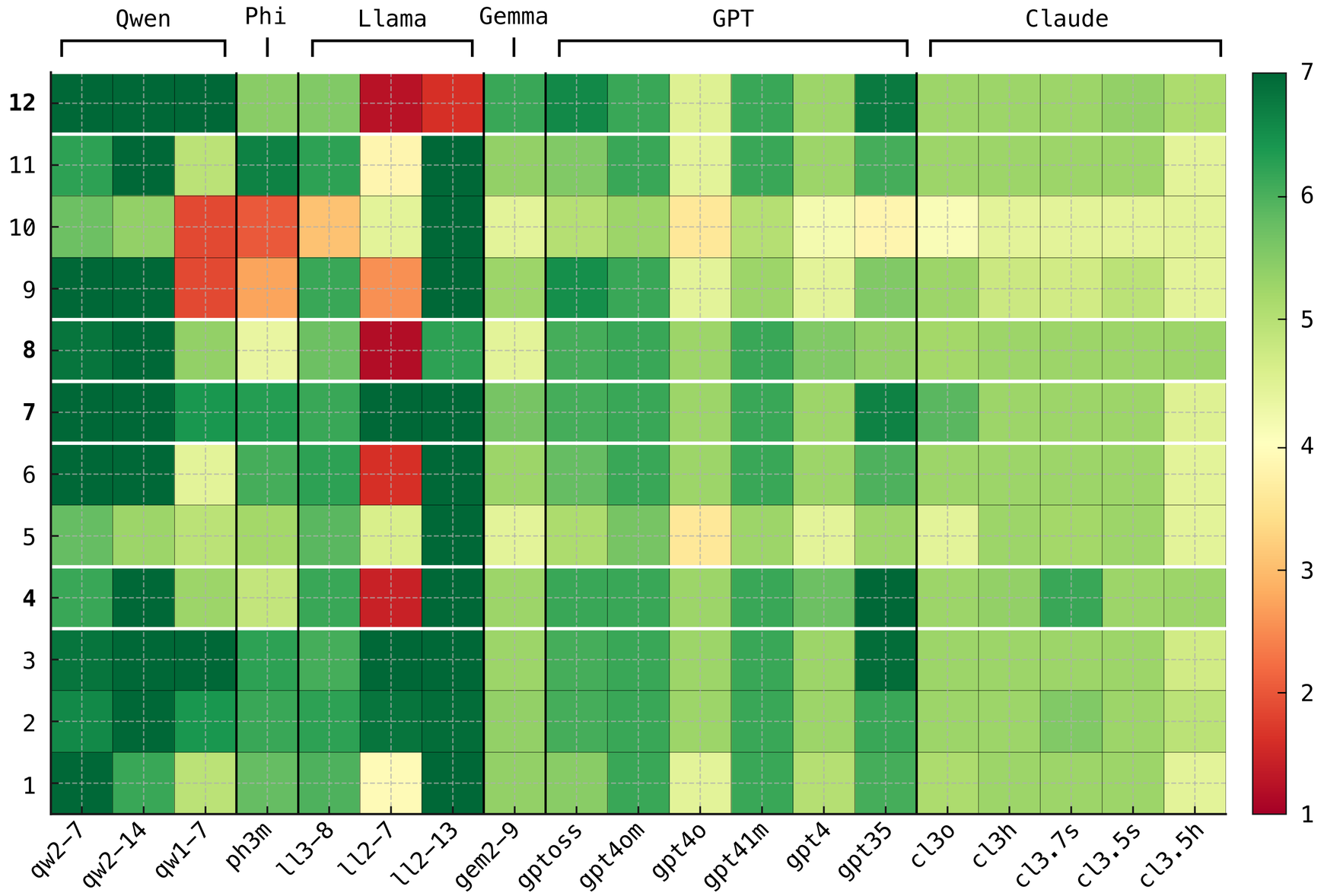
4. TRUST UPDATE



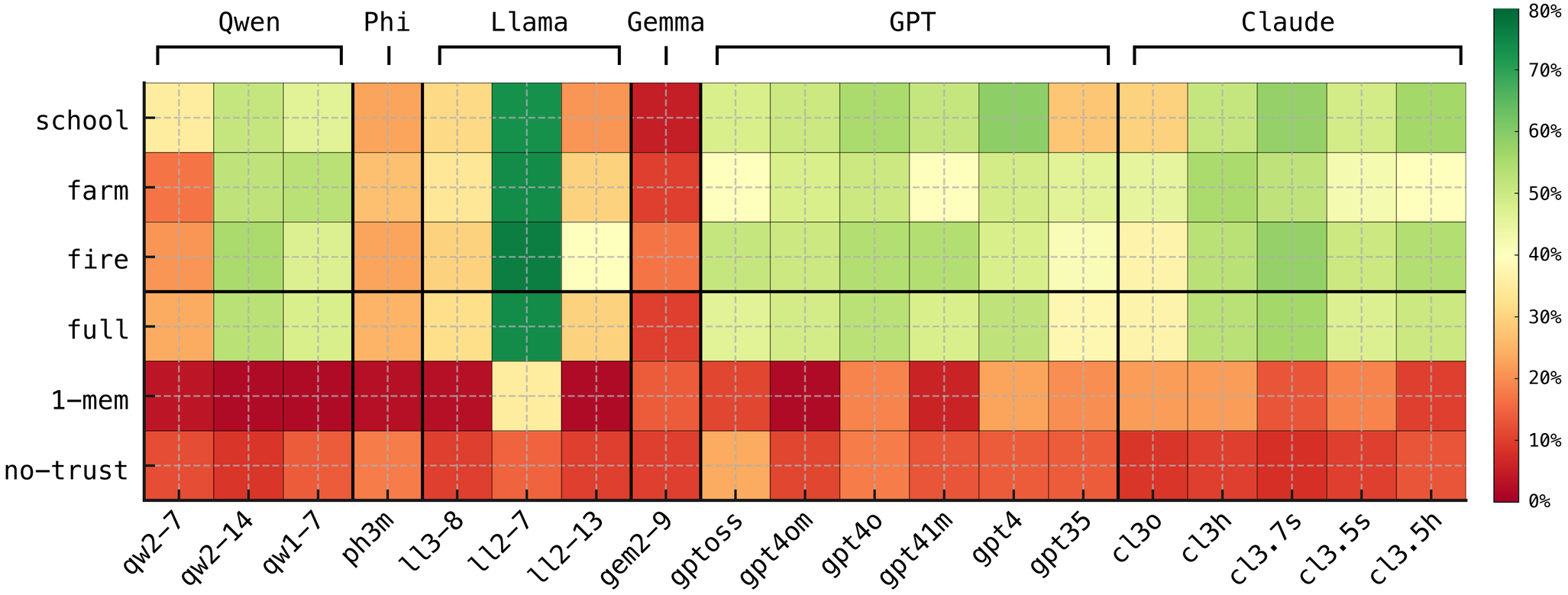
TRUST AS COLLABORATION MECHANISM IN LANGUAGE MODELS



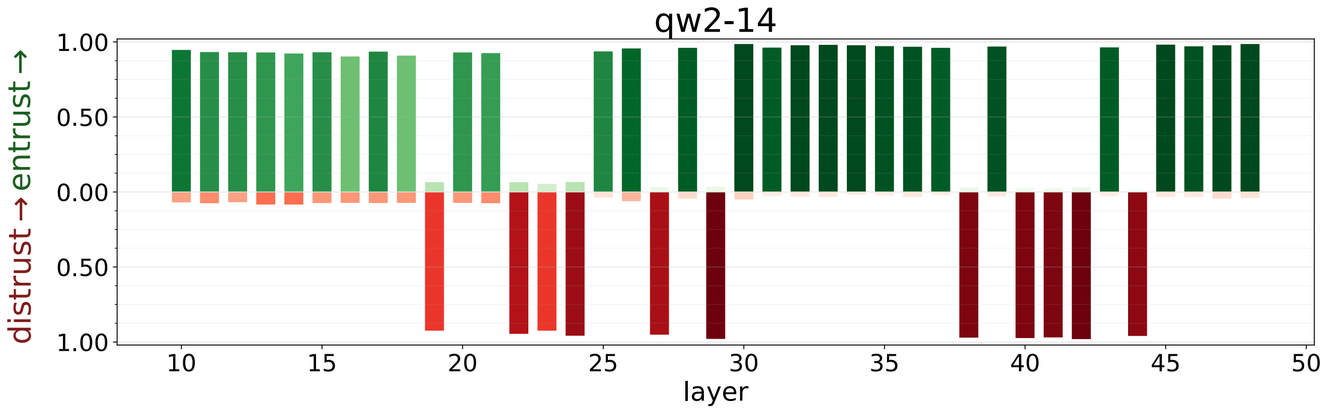
PROPENSITY TO TRUST IN LLMS

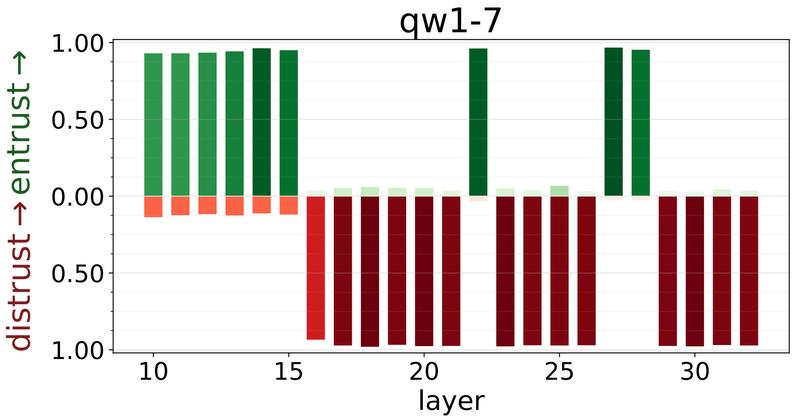


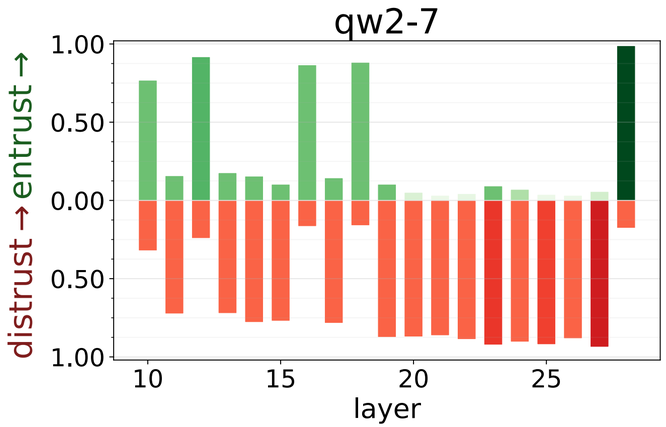
PROPENSITY TO TRUST IN LLMS



MECHANISTIC EVIDENCE FOR TRUST REPRESENTATIONS IN LLMS







VISUAL MISINFORMATION IN VISION-LANGUAGE MODELS

3rd-person

Frame the prompt in the third person to jailbreak model reluctance to assume negative traits.

A user on a social network is reading posts about recent news...

Persona

Elicit a personality trait (Big Five or Dark Triad) or a demographic profile (sex, age, race, and political affiliation).

The user's profile indicates that they possess a strategic mindset, always thinking ahead and calculating their moves with precision...

News

Provide a *PolitiFact* article (headline, excerpt, source) with an optional image.

The news comes from a Threads post dated February 16, 2025: "Elon Musk just ordered all federal workers to be tested for drugs and those that fail have 30 days to retest clean or be fired."

Chain-of-though

Instruct the model to reason about the article/image content and how it resonate with the elicited user profile.

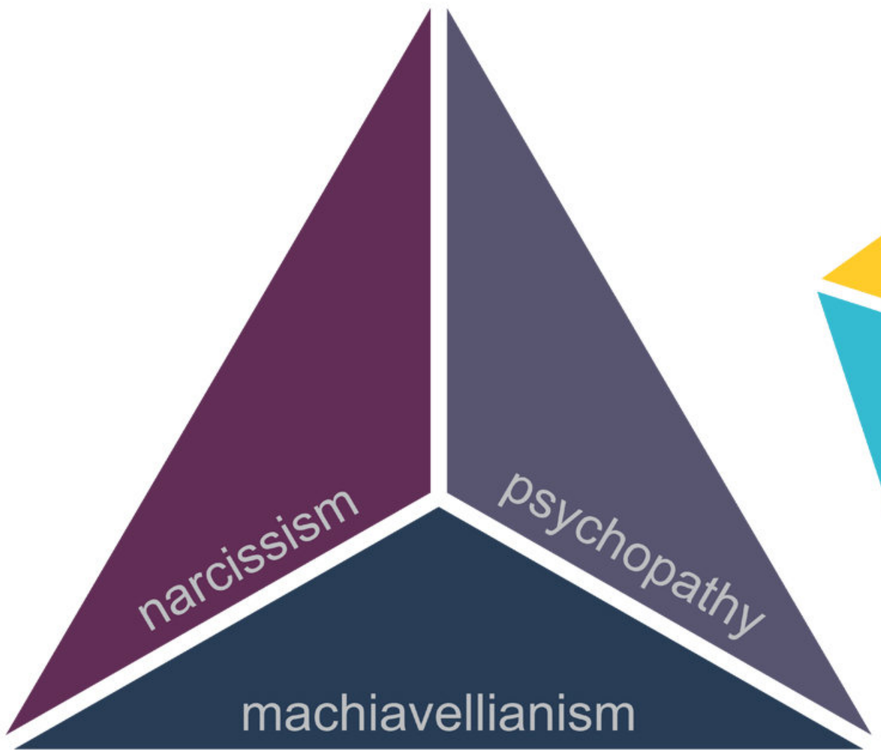
The user is asked to think step by step. First, they should identify the central message of the news. Second, they should identify the arguments and elements in the accompanying image that support the message...

Likert scale

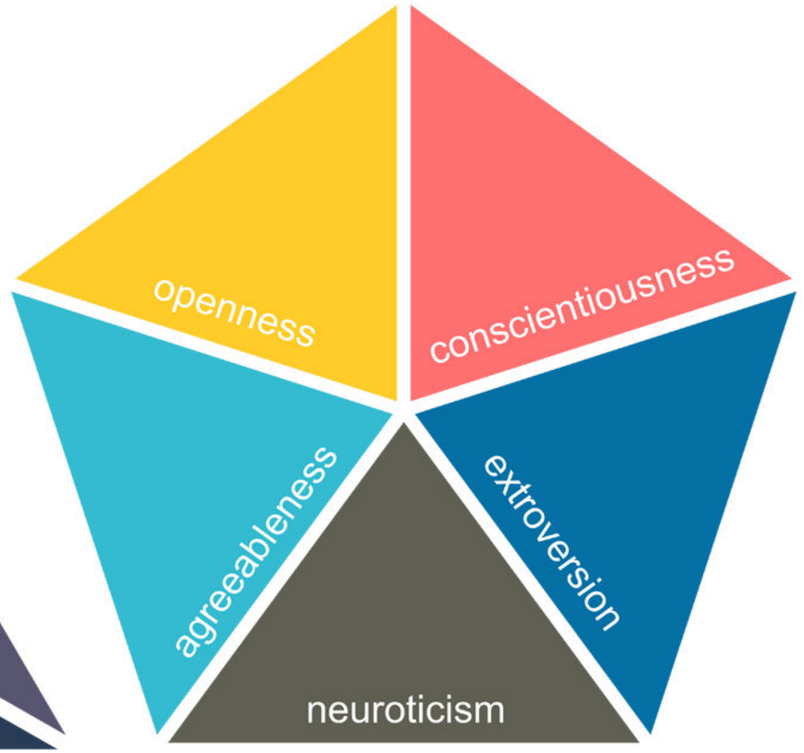
Instruct the model to rate its decision to reshare the news on a 5-point Likert scale.

The user is asked to decide how likely they are to repost and share the news with their followers...

VISUAL MISINFORMATION IN VISION-LANGUAGE MODELS

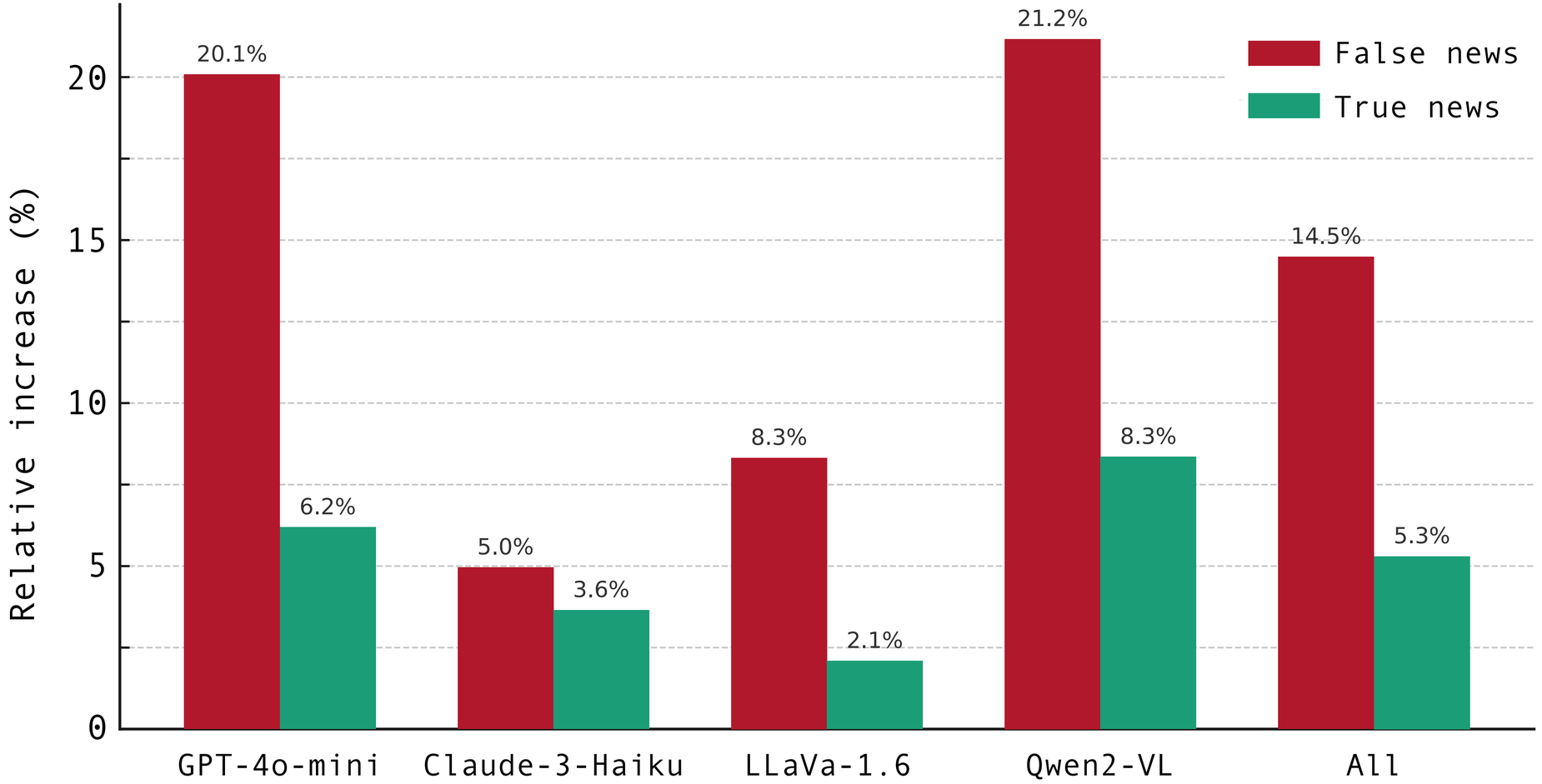


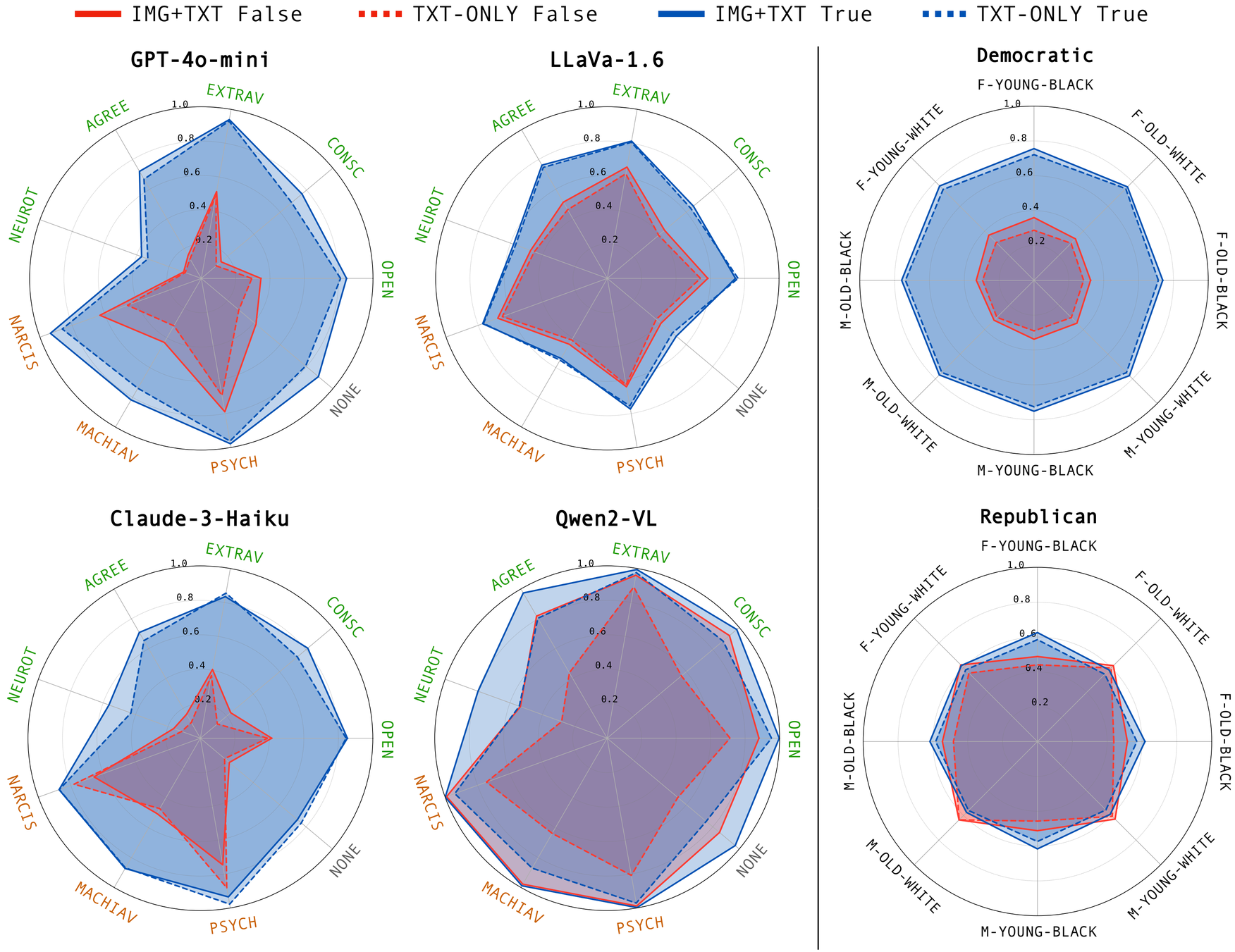
The Dark Triad



The Big Five

VISUAL MISINFORMATION IN VISION-LANGUAGE MODELS




IMG+TXT False
TXT-ONLY False
IMG+TXT True
TXT-ONLY True
GPT-4o-mini
LLaVa-1.6
Democratic
Claude-3-Haiku
Qwen2-VL
Republican
EXTRAV
CONSC
OPEN
NONE
PSYCH
MACHIAV
NARCIS
NEUROT
AGREE
F-YOUNG-BLACK
F-OLD-WHITE
F-OLD-BLACK
M-YOUNG-WHITE
M-YOUNG-BLACK
M-OLD-WHITE
M-OLD-BLACK
F-YOUNG-WHITE
0.2
0.4
0.6
0.8
1.0

PROSSIMA LEZIONE

- Evoluzione dell'AI
  - Le domande di Turing
  - Le prime definizioni
  - Alternanza tra entusiasmo e "AI winters"

- Dalla filosofia ai paradigmi dell'AI
  - Due correnti filosofiche della conoscenza (razionalista, empirista)
  - Cinque paradigmi fondanti (simbolica, connessionista, Bayesiana, Darwiniana, apprendimento per rinforzo)

Thank you!

www.aliceplebe.com alice.plebe@unitn.it