AI GOVERNANCE INSIDE OUT

——A DECISION PERSPECTIVE

YONG TAO

MAY 2024

AGENDA

Urgency for AI governance

Decision-based AI governance

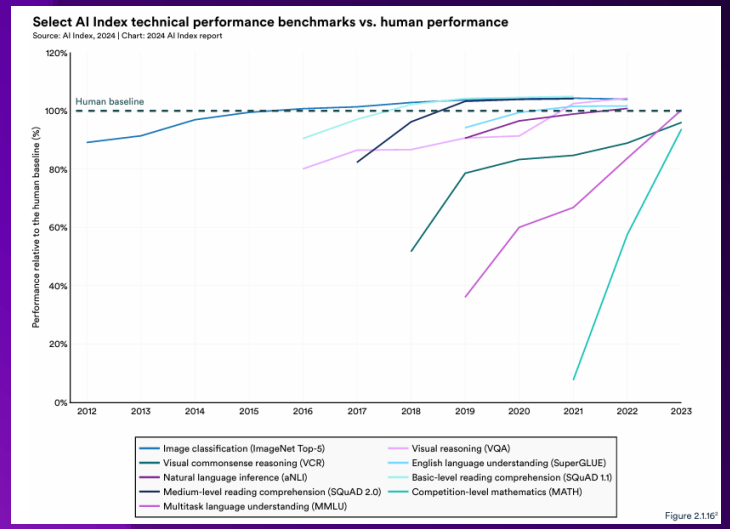
Inside AI decision algorithms & codes

Out to AI developers, testers & users

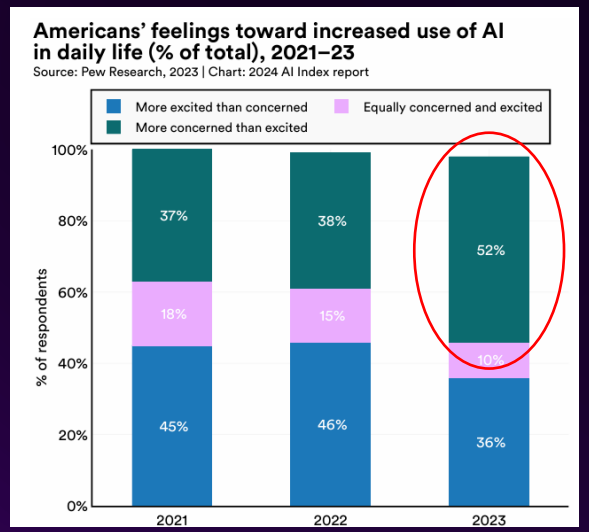
Out further to society

URGENT NEED FOR SYSTEMATIC AI GOVERNANCE

AI is fast surpassing humans in 6 of 9 tasks listed



Sharp Rise in deep public concerns & fear of AI



AI legislation is lagging

- **EU AI Act passed November 2023**
- **US regulations related to AI increases, piece by piece**
- **AI governance Research increases: risk-based approach**

Source: Stanford HAI Artificial Intelligence Index Report 2024

GOALS OF AI GOVERNANCE

Promote AI benefits to society;

Ensure AI acts within boundaries of ethics, laws, & regulations (ELRs);

Maintain human control in case of AI emergency;

Mitigate AI risks.

HOW OTHER TECHNOLOGIES ARE GOVERNED: PASSIVE TOOLS EXTENDING HUMAN HANDS, GOVERNING HUMAN DEVELOPERS & USERS

Governance of other technologies:

- **Passive tools extending human hands (legs)**
- **Decisions are made by developers and users**
- Developers are governed by **ethics**, **laws**, and technology related industry **regulations** (**ELRs**) to provide and guarantee specified performances
- Users are fully responsible for consequences of technology usage according to **ELRs** of society
- Technology is a triple-edged sword: the good, bad, and **accidents**

AGENDA

Urgency for AI governance

Decision-based AI governance

Inside AI algorithms & codes

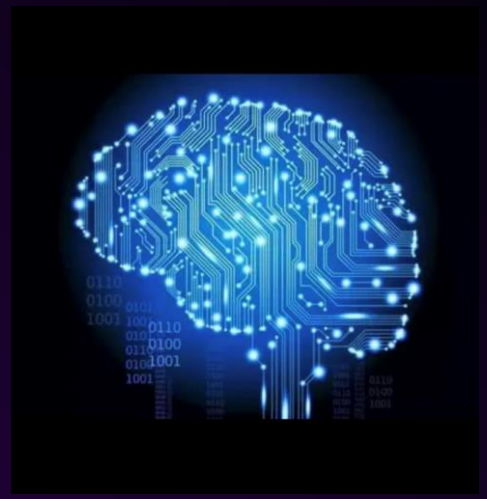
Out to AI developers, testers & users

Out further to society

AI DIFFERS FROM OTHER TECHNOLOGIES; AI SHOULD BE GOVERNED BY DECISIONS IT MAKES

AI differs from other technologies

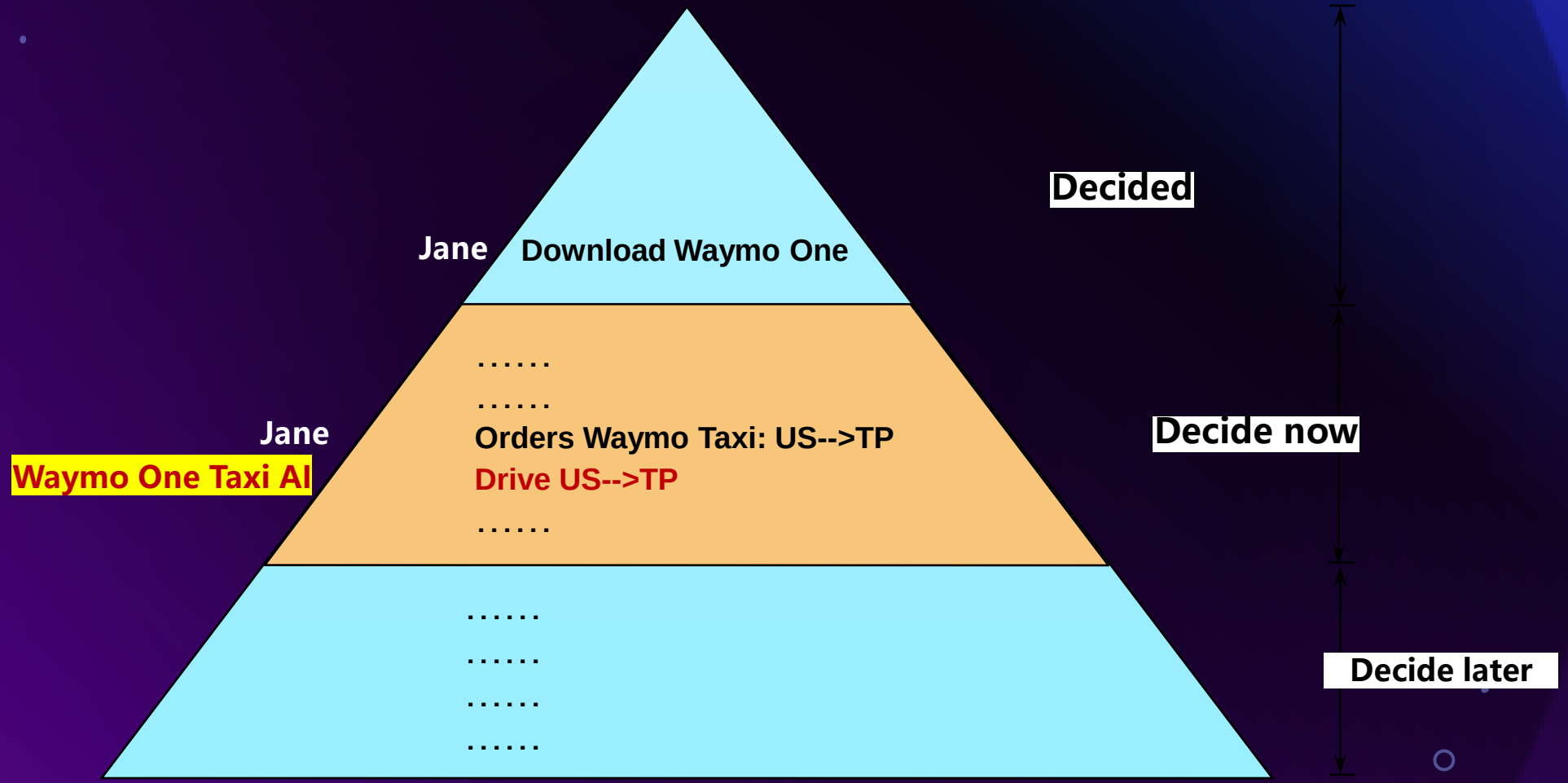
1. **Extensions of the human brain**
2. Helps humans make decisions
3. **AI can proactively make independent decisions**
4. **AI decisions interact with humans and impact our lives**



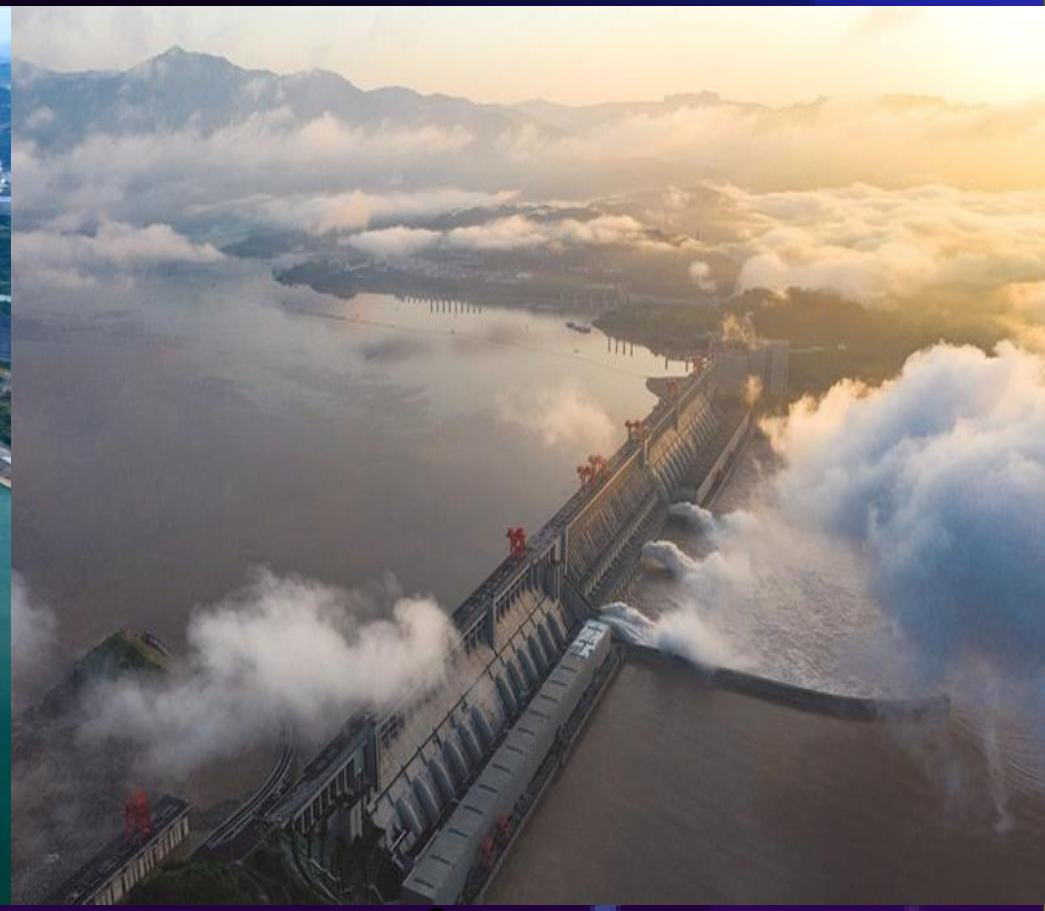
Key questions of AI governance

- **Humans are governed by our decisions and actions through ELRs (Ethics, Laws, Regulations)**
- **AI decisions interact and impact humans, AI should be governed by the same ELRs as humans**

HUMAN-AI DECISION HIERARCHY



VAST RESERVOIR OF INTELLIGENCE VS DECISIONS FLOW OUT OF THE RESERVOIR



WHAT TO GOVERN, AND NOT TO GOVERN?

What NOT to govern

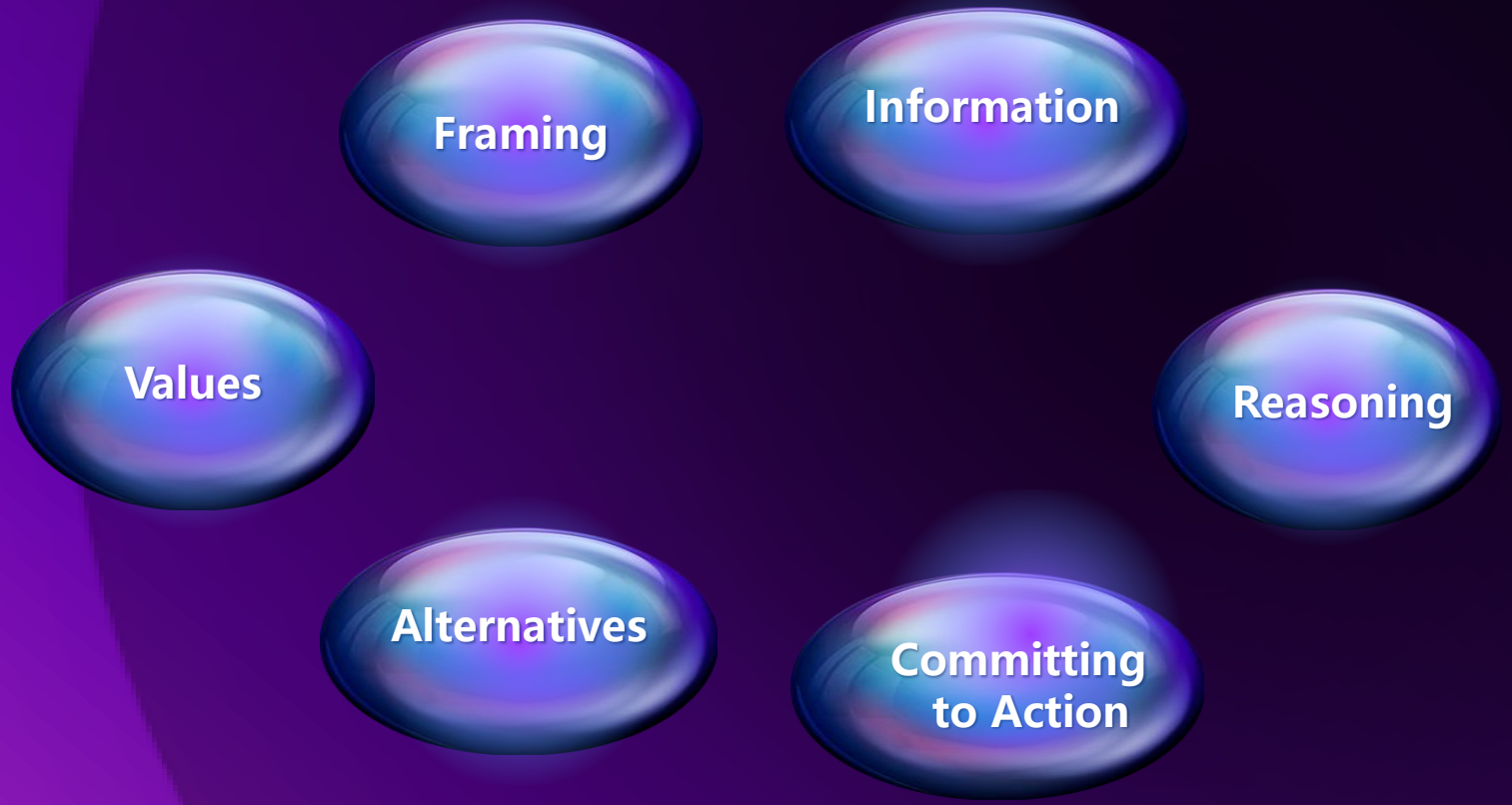
- **Intelligence per se --- latent**
- **Vast reservoir of AI "thoughts"**
- **We do not govern human thoughts**
  - >6000* thoughts/day for a human, only a tiny part end in decision

What to govern

- **AI decisions flowing out of AI reservoir**
- **Better yet, deep down to AI decision process**

* Source: Julie Tseng & Jordan Poppenk, "Brain meta-state transitions demarcate thoughts across task contexts exposing the mental noise of trait neuroticism", NATURE COMMUNICATIONS| (2020) 11:3480 |https://doi.org/10.1038/s41467-020-17255-9|www.nature.com/naturecommunications

GOVERNING AI DECISIONS BETTER BY GOING TO KEY AI DECISION ELEMENTS



AI decision process:

- Open AI black box
- More leverage points
- Better AI decisions
- More AI Explainability

AI GOVERNANCE: A MINIMUM SET OF NEW RULES (MNR), HARMONIZE WITH SOCIETAL ETHICS, LAWS, AND REGULATIONS (ELR)

Minimum set of New AI Rules (MNRs)

- Enable ELR adaptation to AI
- MNRs allow AI transparency
- Enable human action & resource control in emergency
- MNRs Calm deep public fear of AI

Ethics, Laws, and Regulations (ELRs)

- Existing ELRs governing humans
- No need to reinvent ELRs for AI
- Need MNRs to adapt ELRs for AI
- MNRs only complement ELRs, not replacing them

AGENDA

Urgency for AI governance

Decision-based AI governance

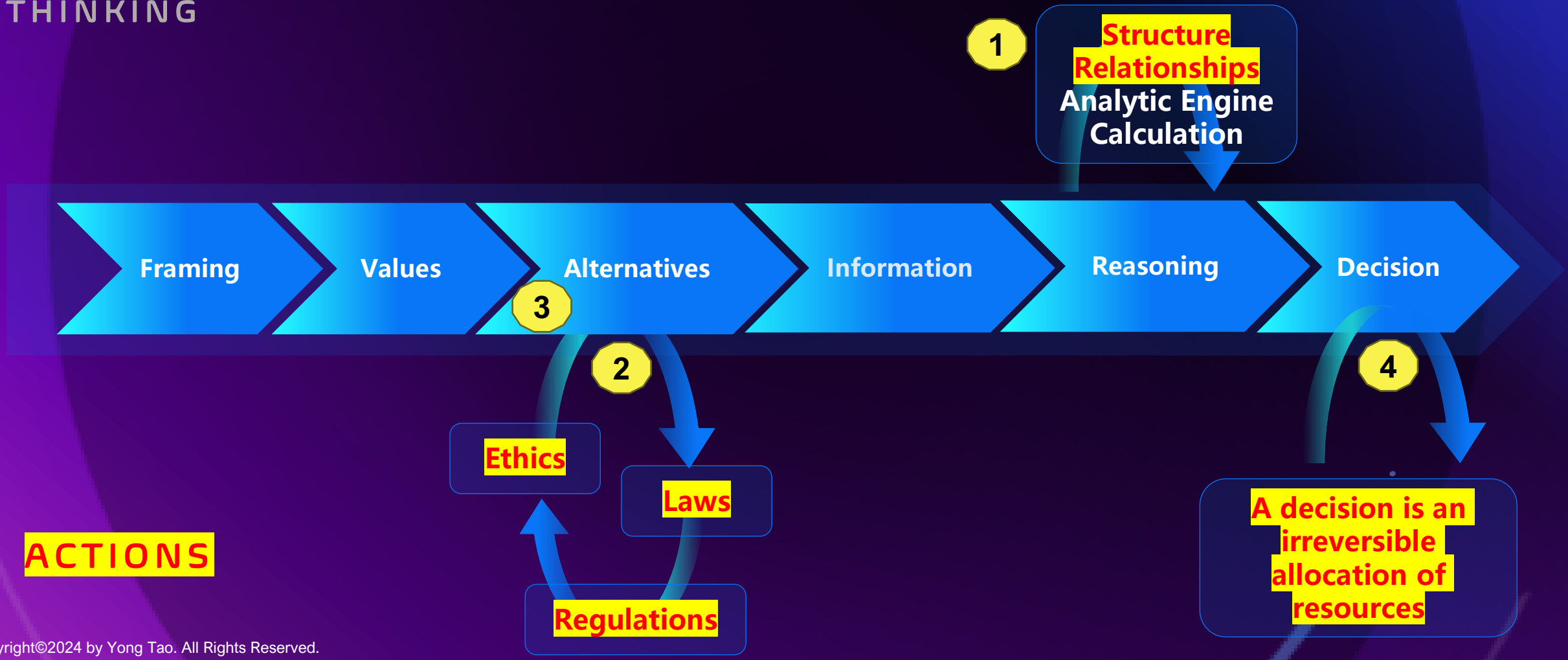
Inside AI decision algorithms & codes

Out to AI developers, testers & users

Out further to society

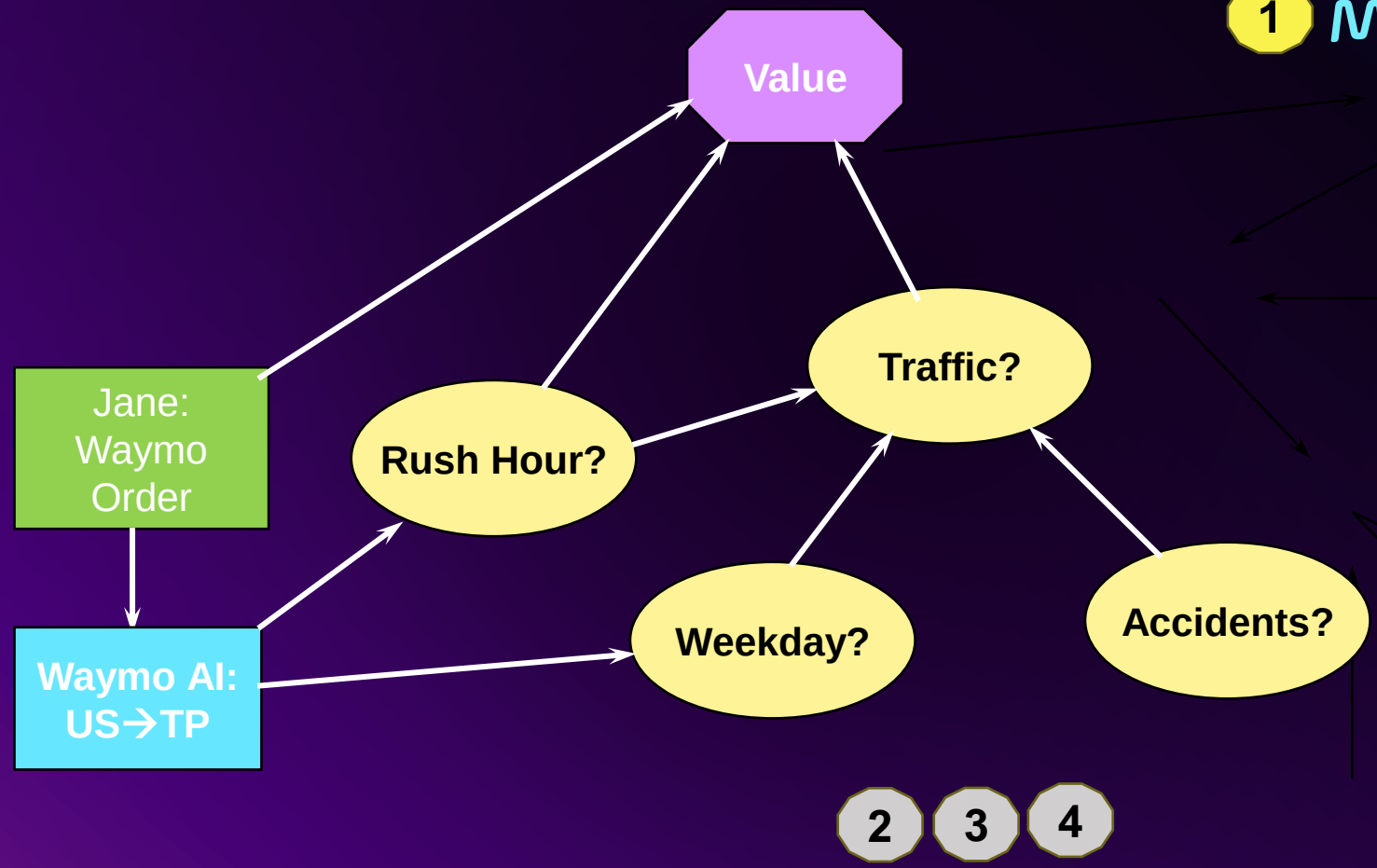
GOVERNING AI AT KEY LEVERAGE POINTS INSIDE DECISION ELEMENTS FOCUSING ON ACTION ALTERNATIVES

THINKING



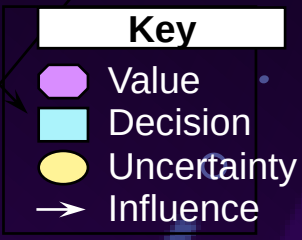
ACTIONS

MNR 1: (INSIDE AI) STRUCTURAL TRANSPARENCY AND MINIMUM GRANULARITY



1 MNR 1 (Inside AI):

- Decompose AI decision into key elements
- Isolate misinformation

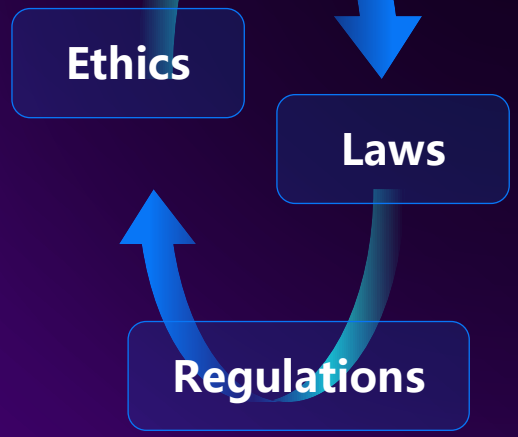


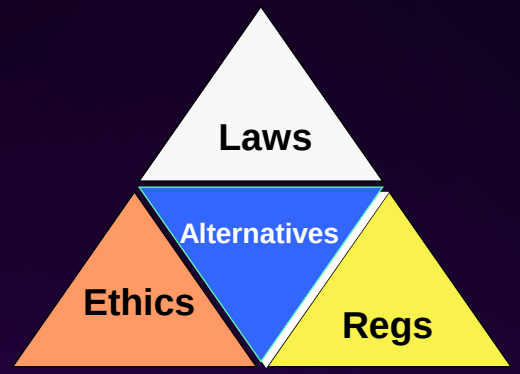
MNR 2: EACH ACTION ALTERNATIVE AI EVALUATES MUST SATISFY EXISTING SOCIETAL ELRS

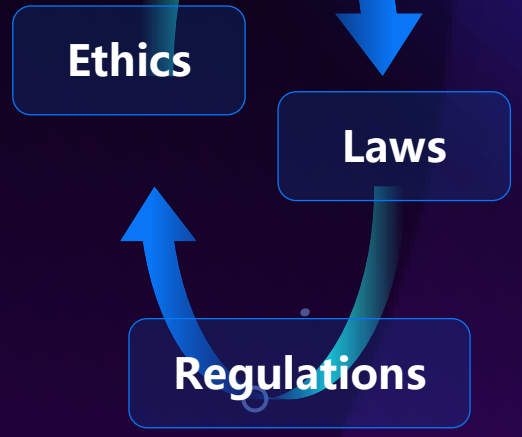
MNR 2 (Inside AI): 2

Level playfield as humans (ELRs)

Biggest area of public trust deficits





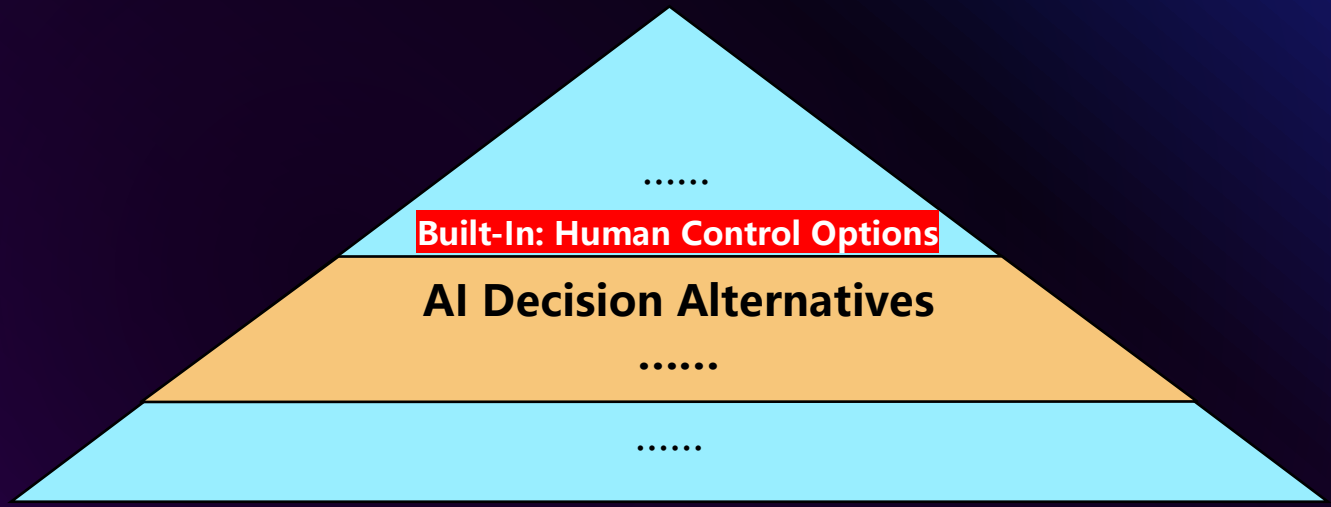


AI FINAL DECISION ACTIONS MUST SATISFY ELR TESTS

MNR 3: BUILT IN HUMAN INTERVENTION OPTIONS AT A LEVEL HIGHER THAN AI IN THE HUMAN-AI DECISION HIERARCHY

MNR 3 (Inside AI): 3

- Gives user control of AI
- User can pre-authorize
- Human control (developer or others) at request of user in existential threat, or catastrophic accidents



Framing → Values → Alternatives → Information → Reasoning → Decision

MNR 4: HUMAN AUTHORIZATION IS BUILT IN BEFORE ALLOCATING USER'S RESOURCES TO EXECUTE AI DECISION

4 MNR 4 (Inside AI):

- AI decision allocates user resources
- User authorization must be built in before decision execution
- User can pre-authorize

Framing → Values → Alternatives → Information → Reasoning → **Decision**

AGENDA

Urgency for AI governance

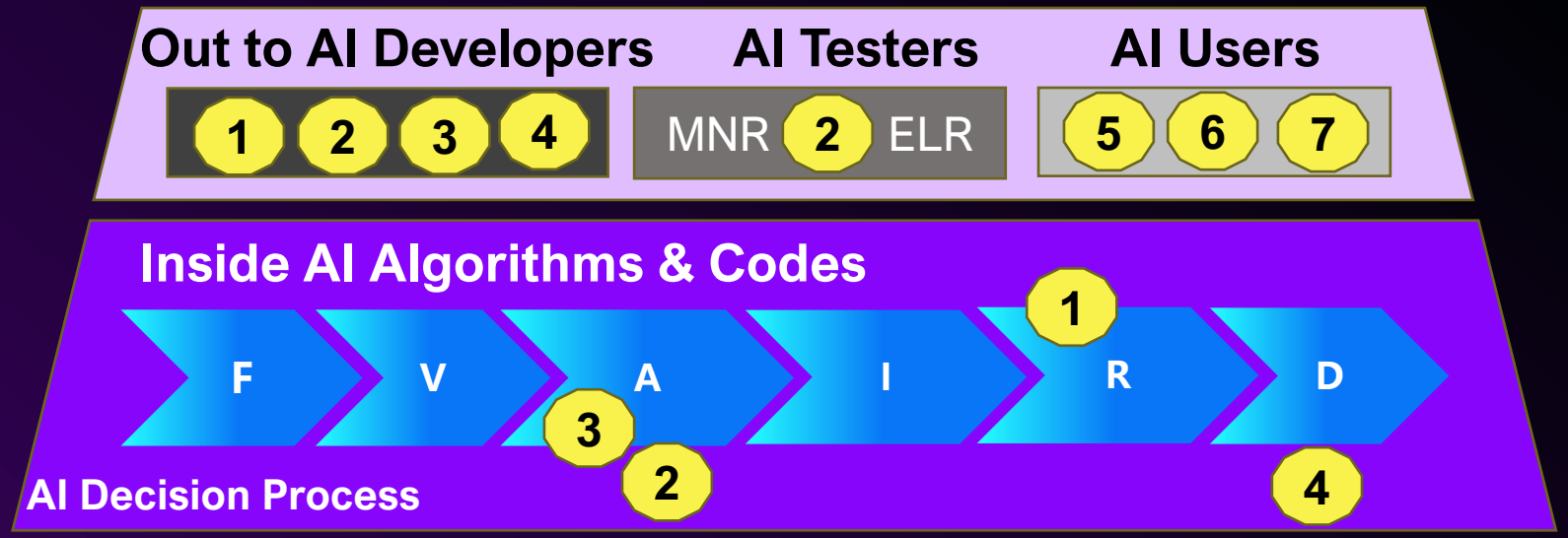
Decision-based AI governance

Inside AI algorithms & codes

Out to AI developers, testers & users

Out further to society

OUTSIDE AI SYSTEMS, AI DEVELOPERS, TESTERS, AND USERS' RESPONSIBILITIES AND OBLIGATIONS ARE CRUCIAL FOR EFFECTIVE AI GOVERNANCE



MNR 5: USERS HAVE RIGHTS AND OBLIGATION TO INTERVENE AI DECISION ACTIONS

MNR 5 (OUT TO USERS): 5

IN CASE OF FATAL ACCIDENTS, CATASTROPHE, EXISTENTIAL THREATS, AI USERS ARE OBLIGATED TO INTERVENE THROUGH BUILT IN OPTIONS 3

MNR 6: HUMAN USERS HAVE THE RIGHTS AND OBLIGATION TO CUT OFF RESOURCES FOR AI DECISION EXECUTION

MNR 6 (Out to users): 6

Cut off resources for AI decision execution

Stop authorization of allocating, and cutting off user's resources to execute AI decisions as user see fit

If catastrophic accidents or existential threats occur, user must cut off resources or may request developer or other humans to cut off resources 4

MNR 7: USERS HAVE RIGHTS, OBLIGATION TO CUT OFF ENERGY SUPPLY TO AI

MNR 7 (Out to users): 7

Cut off energy supply to AI if needed

- Stop AI as user see fits
- If catastrophic accidents or existential threats occur, user must cut off energy supply to AI or request developer or other humans to do so

AGENDA

Urgency for AI governance

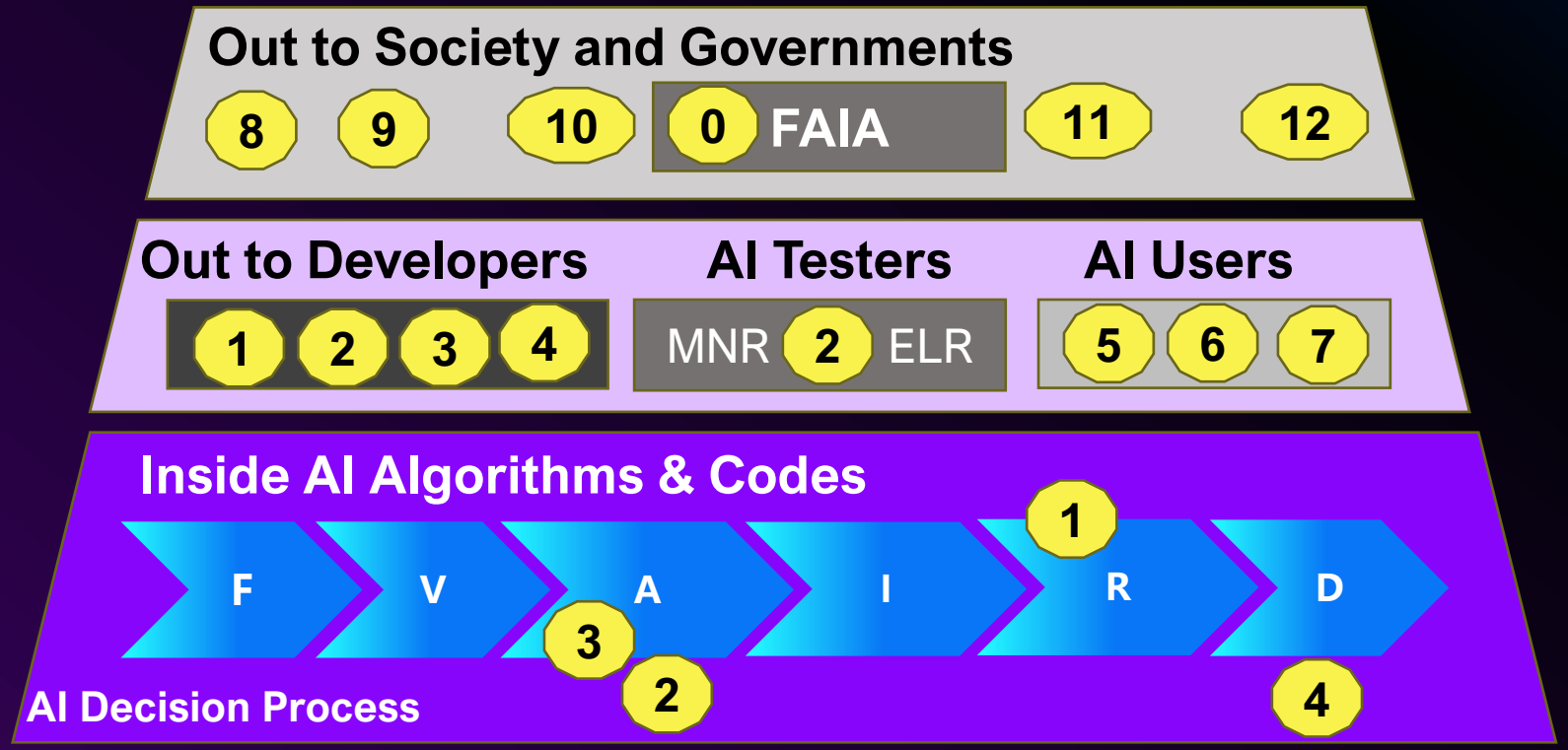
Decision-based AI governance

Inside AI algorithms & codes

Out to AI developers, testers & users

Out further to society

STRONG GOVERNMENTS AND SOCIETY LEVEL AI GOVERNANCE INFRASTRUCTURE IS NEEDED



MNR 8: AI HAS ITS OWN CLASS OF IDENTITY, DIFFERENT FROM HUMANS

MNR 8 (Out to society): 8

AI systems, AI products, AGIs will have a different class of Identity from humans.

Carbon based organic humans build silicon based inorganic AI as our helpers, agents, disciples, not as an equal, even though AI capability already surpasses humans in some tasks. 9

MNR 9: AI HAS NO RIGHTS TO OWN PROPERTIES AND RESOURCES

MNR 9 (Out to society): 9

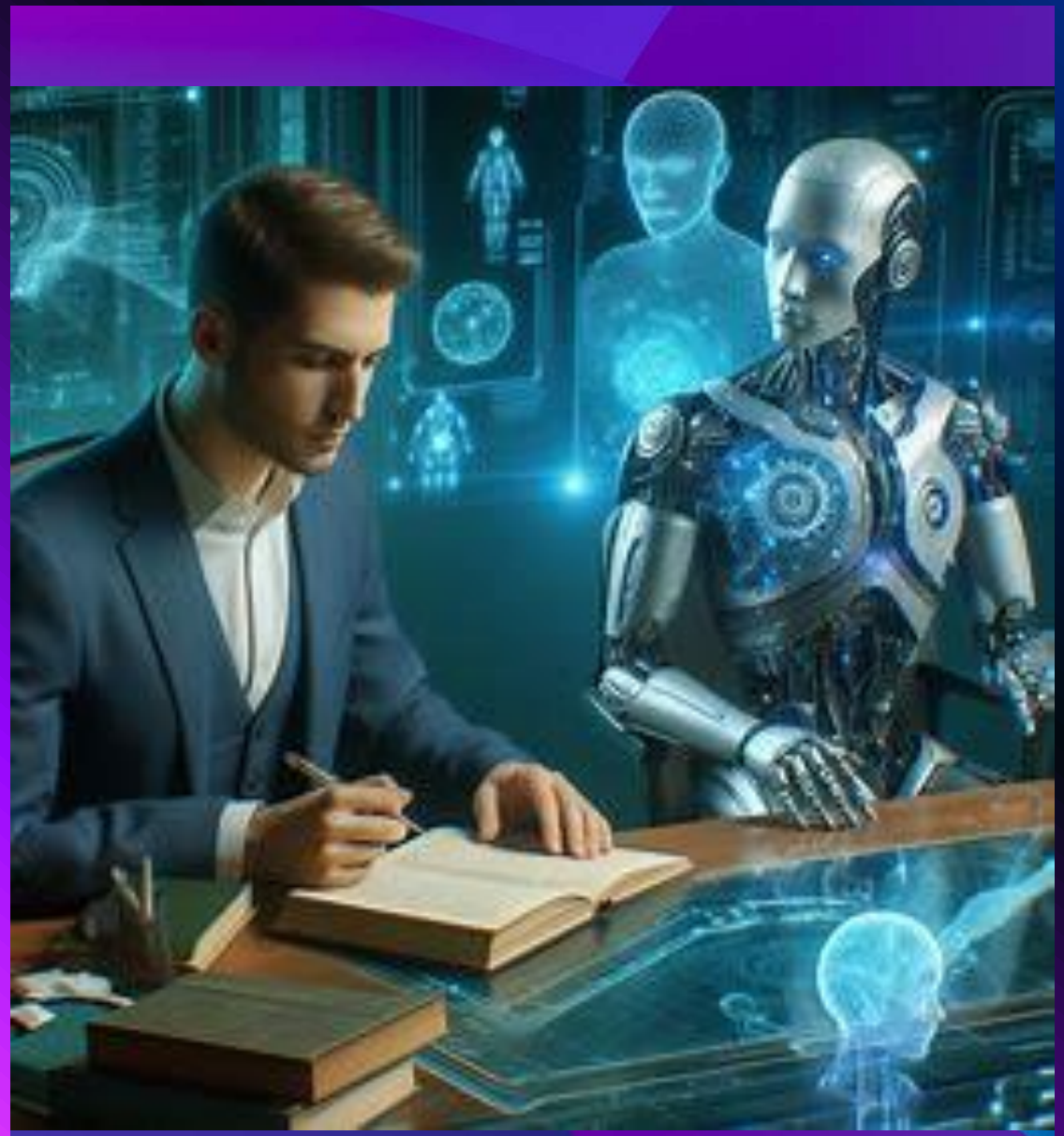
AI systems, AI products, AGIs do not have the rights and privileges to own properties and resources. 7

AI can only obtain authorization from humans to allocate resources for AI decision execution. 4 6

MNR 10: AI HAS NO RIGHT TO PUBLISH EXPRESSIONS, PAPERS, BOOKS INDEPENDENTLY

MNR 10 (OUT TO SOCIETY): 10

GENERATIVE AI CAN WRITE BUT CANNOT TAKE RESPONSIBILITY FOR ITS EXPRESSIONS, THUS NO RIGHTS TO PUBLISH INDEPENDENT OF HUMAN



MNR 11: HUMAN USERS MUST LIST AI AS CO-AUTHOR IF AI GENERATES PARTS OF A PUBLICATION

11

MNR 11 (Out to society):

AI systems, AI products, AGIs must be identified and labeled as co-authors for the publication AI generates.

Human author(s) takes ultimate responsibility for the publication.

MNR 12: SEPARATION OF AI (DECISIONS) FROM POWERFUL PHYSICAL TOOLS

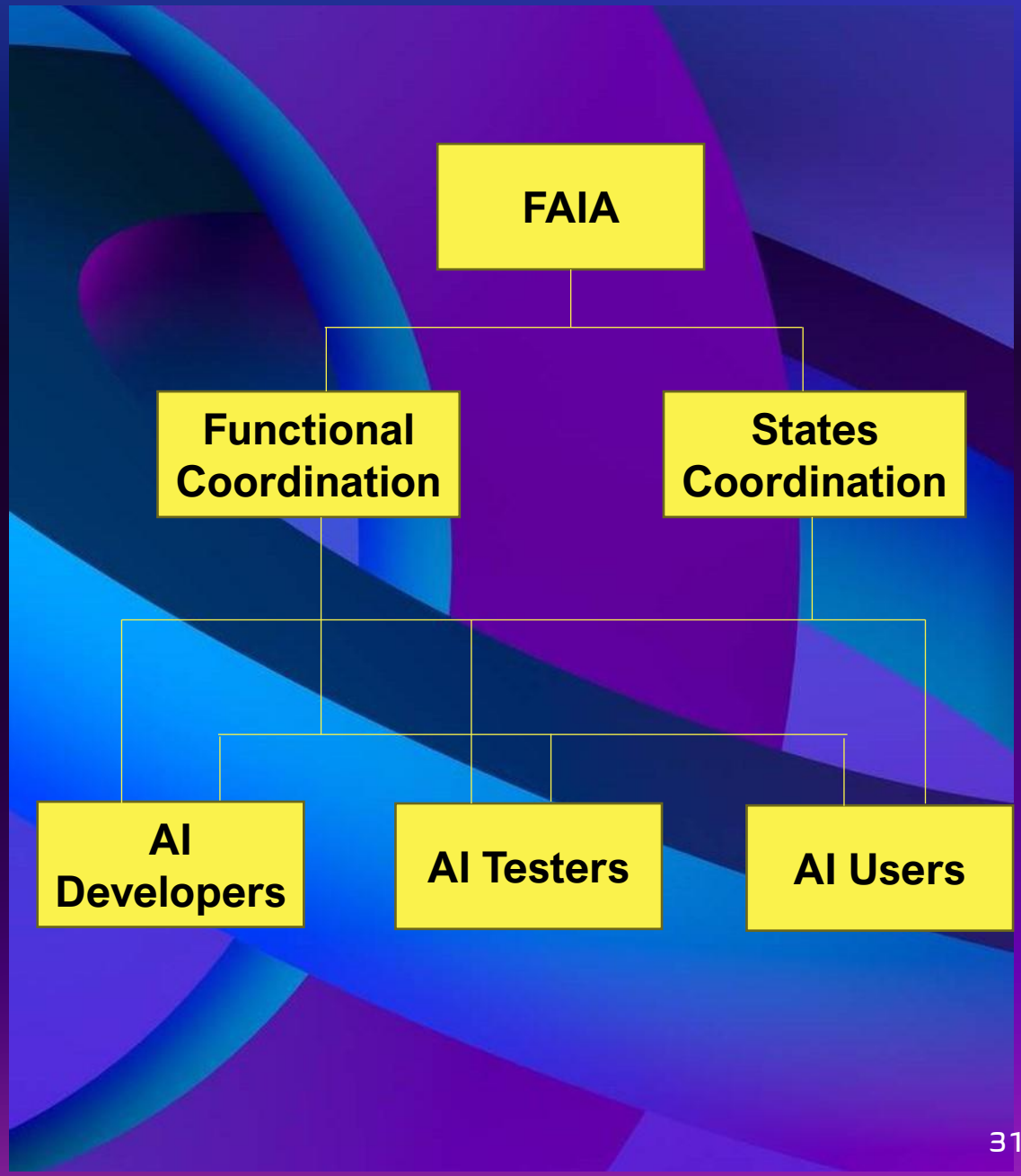
12 **MNR 12 (Out to society):**

AI systems, AI products, AGIs must be separated from powerful physical tools and weapons, with human intervention in between.

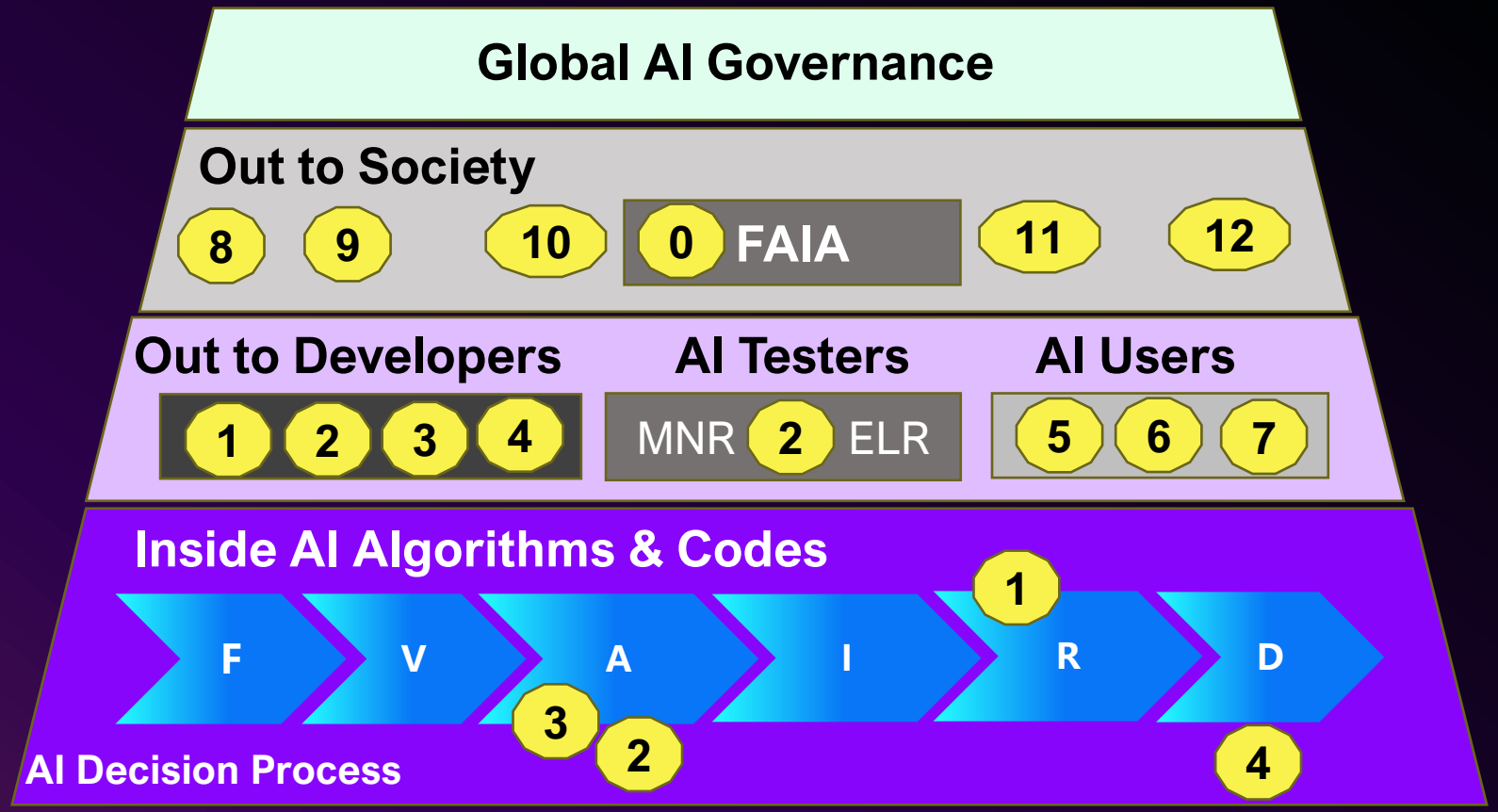
This introduces subjectivity as to what is powerful. To be conservative, this must be done to give public a peace of mind.

MNR 0: 0 ESTABLISH FAIA (FEDERAL AI ADMINISTRATION)

FAIA IS NEEDED TO OVERSEE AI SYSTEMS MAKING DECISIONS INTERACTING WITH HUMANS AND IMPACT OUR LIVES



GLOBAL AI GOVERNANCE



Global AI Treaty on:

- AI Weapons of Mass Destruction
- AI Existential Threats to Humanity
- Timely sharing of best practices of national AI Governance

GOVERNING AI DECISIONS, NOT AI

Yong Tao

+1 650-704-9852

yongtao@cloudstonevc.com

www.cloudstonevc.com