# AI for Network Automation

Laurent Ciavaglia

IEEE NOMS Keynote

May 2024

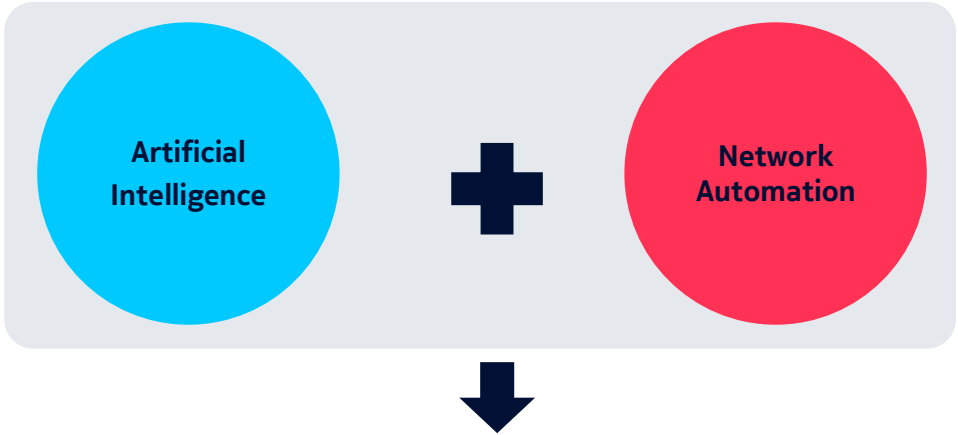
Credits and thanks to my colleagues for the materials and the inspiration

For any complaints or difficult questions, please contact them

(I can give you the names…) 

# Scope




Advanced integration and embedding capabilities needed to maximize the use and efficiency of **AI-assisted and AI-enabled** automation

# High potential areas

## Network planning

e.g. accelerate planning decisions by transferring knowledge/models learned in other operational domains

## Service/slice provisioning

e.g. service/slice creation and deployment based on automated descriptor checking, fetching and completion

## Operator experience

e.g. transform the operator experience by taking advantage of machine learning and reasoning in adapting control and supervision interactions

## Security

e.g. AI-powered threat detection and mitigation, data integrity assurance

## Network optimization

e.g. leverage real-time analytics to dynamically select optimal routes and steer traffic in multi-technology and cross-domain/cross-layer topologies

## Service/slice and network assurance

e.g. reduce high volume of incidents to meaningful situations (e.g. outage and congestion prediction, SLA assurance…) by predicting, detecting and correlating events

## Intent fulfillment

e.g. learn what actions are more efficient and impactful to realize intents in given contexts with self-evaluation and self-measurement capabilities

## And others such as…

- Testing
- What-if scenarios
- Tactical decision-making…

# Main drivers…

## Technology

Major breakthroughs. (backpropagation algorithms, convolutional neural networks, generative adversarial networks, high-dimensional word embeddings…)

Continuous advances in techniques and algorithms (e.g. graph and adversarial neural nets)

Ethical and regulatory considerations (e.g. differential privacy, explainable AI) becomes integral parts of the technology development

## Data

Transformation from CSP* to DSP*, in-line with holistic digitization

Data collection platforms becomes intelligent, agile and with expanded application space

Increased use of data lakes, data meshes

Techniques and tools available for synthetic data generation

## Compute

Moore's law and continuous evolution of HW specialization e.g. CPU, GPU, TFU, ASICS…

Compute continuum: spanning from devices/sensors, through the edge to core DC

## Tools

Maturing ecosystem of platforms, languages and libraries (TensorFlow, SageMaker, scikit-learn, PyTorch, Keras…)

Progress and availability of pipeline automation (e.g. AutoML)

## Business

Embracing the industry evolution and new opportunities with the adoption of AI technologies and practices across CSP and NPN scenarios

Web-scale actors driving the main AI technology building blocks and steering their development

Attracting DSP data and increasing collaborations between DSP and web-scale partners – with/without telco vendors

# … and gaps

## AI mainstream is myopic about other AI realms

Heavily focused on Deep Learning, for human perception tasks

Semantic learning, knowledge representation/embedding, common sense reasoning… still minority but highly needed to support the full cycle of network automation

## AI-based solutions working in isolation and case-tailored

Necessary coordination between multiple AI solutions for coherent decisions

Generalization and automation methodologies are needed to streamline AI pipelines

Radical re-thinking how to apply AI for the design and operation of network automation; go beyond simple replication of the usual problem-solution design process with AI

Application of AI technologies in NPN scenarios less investigated than for CSP

## Limited use or reliance on standards

Still in the Infancy age of Networks AI and Networks AI standards

Standards needed to allow value creation with AI in multi-vendor and open solutions

No or limited interoperability and re-usability of de-facto standards driven by software development

AI technologies are developing and expanding fast; however, some gaps must be overcome to fully benefit from AI technologies potential

# Challenges

## 1

## Networks are hard AI problems

- Networks are distributed, dynamic, heterogeneous, and encrypted
- Complexity in data sets and complexity in the algorithms that deal with them
- Network data are characterized by high-dimensional spaces
- Network data is heterogeneous and diverse
- Access to network data is difficult, lack of reference (labeled) data sets

# Challenges

## 2

## Confidence in AI

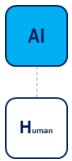
- AI technologies introduce new challenges and specificities such as
  - Performance degradation (re-training)
  - Data sensitivity: bias, adversarial inputs, governance (security, privacy)
  - Shift from deterministic/provable algorithms to stochastic/statistical paradigm
  - Explainability, accountability
- Need for a comprehensive framework to increase confidence in the use of AI

# Challenges

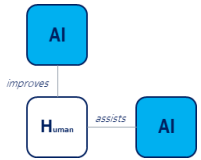
## 3

## Human – AI interaction

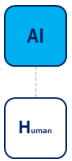
- Various levels of intelligence and support to human operator from assisted to empowered
  - Assist or extend human operator in service design, parametrization and operation
  - Adaptive intelligent automation based on dynamic levels of autonomy and supervision
- Different roles and interactions between the AI-based system and human operators
  - Introducing additional standardization requirements
  - Evolution of human operator's skills e.g. tune AI system, different forms of policies, and AI output interpretation
  - Adaptation of the AI output representation with the relevant domain knowledge and system abstractions




**Independency**
AI performs beyond immediate human operator control




**Enablement**
AI work with human operator in close cooperation




**Replacement**
AI does same task as human operator

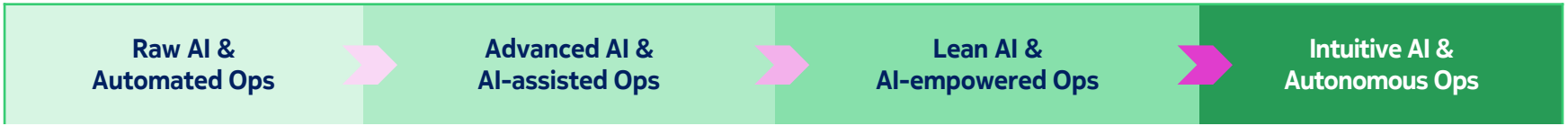
# Challenges

## 4

## AI diversity

- AI techniques have different requirements and standardization needs
- Overall design integrating and leveraging individual AI properties
- Coordination between distributed AI applications to ensure coherent e2e operation
- Additional (AI-specific) software diversity

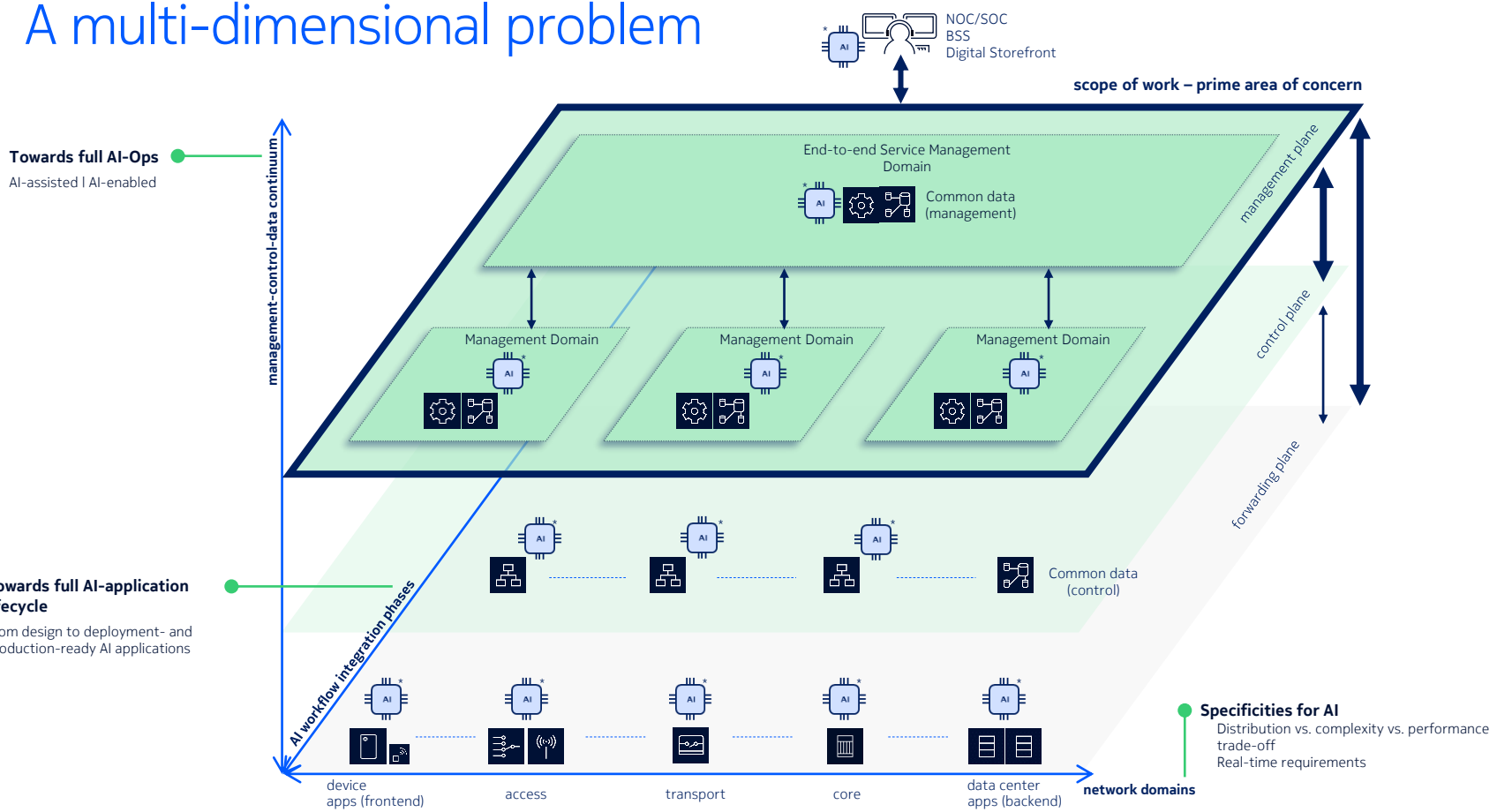
# Joint evolution of AI and Ops



# Joint evolution of AI and Ops

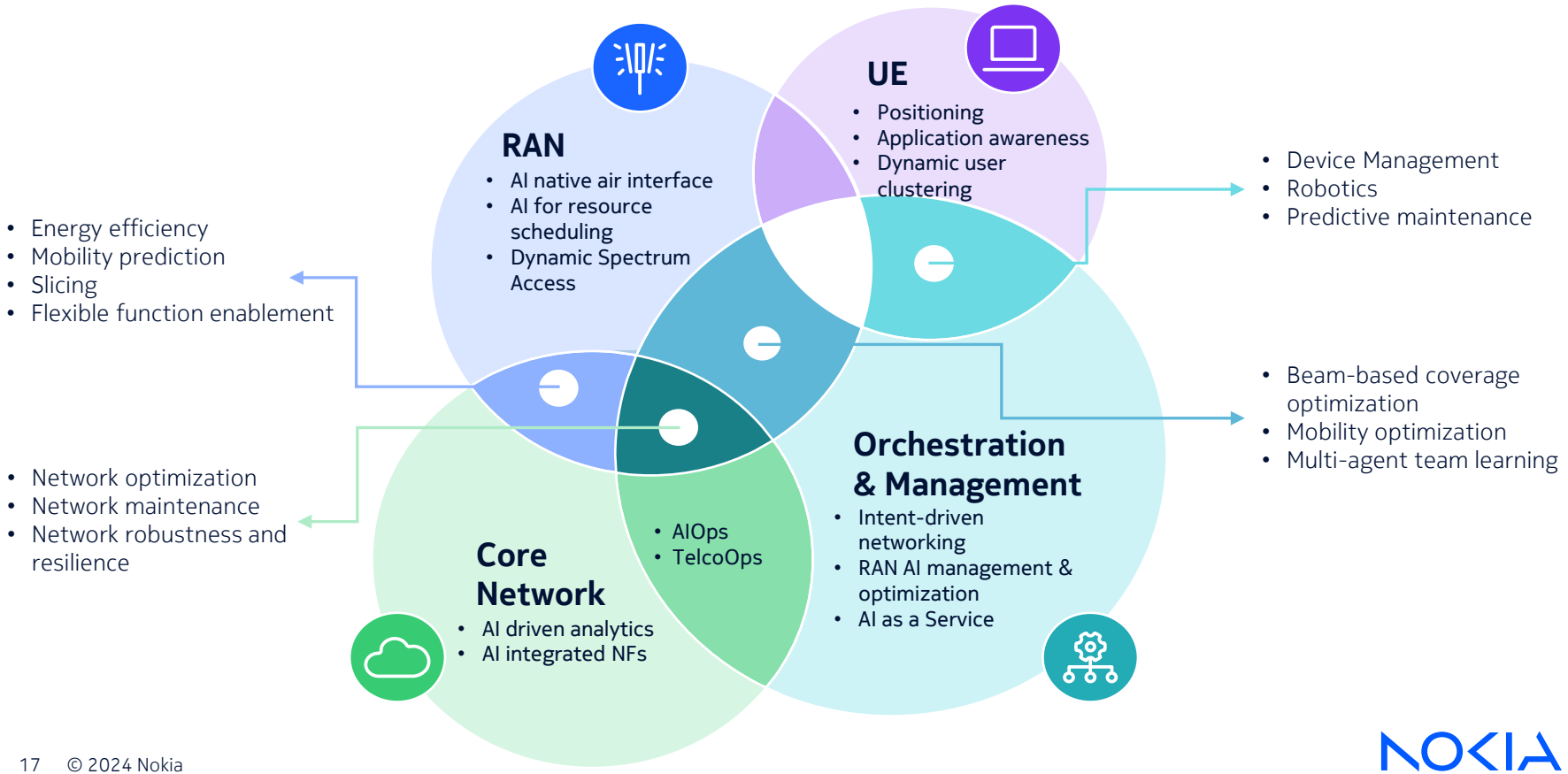
| | Raw AI & Automated Ops | Advanced AI & AI-assisted Ops | Lean AI & AI-empowered Ops | Intuitive AI & Autonomous Ops |
|---|---|---|---|---|
| AI & Data | Limited view and use of AI potential<br>Big dumb data | More diversified, network-adapted AI techniques<br>Smarter data | Broad set of AI techniques for N&S environment<br>Intelligent data | Zero-touch AI-Ops<br>Machine Reasoning<br>Symbiotic Human-AI interaction<br>Mission autonomy<br>Transparent, trusted, open AI<br>Reliable, robust and distributed AI |
| Scale & Adoption | Use case-driven<br>Isolated, small scale solutions with limited re-use | Cross use cases<br>Large scale application and penetration of AI-based N&S automation solutions | "AI-as-a-Service"<br>Full scale deployment and applicability of AI-enabled, plug-n-play solutions | |
| Practice & Integration | Retrofit ML technologies for N&S automation<br>Manually-intensive integration | AI know-how is leveraged for N&S automation<br>Semi-automated design and integration | Designed with AI<br>Seamless design and integration | |
| Confidence & Security | Controlled autonomy and confined in scope<br>No AI-specific security measures | Towards operation autonomy<br>Trust framework safeguards AI-based solutions<br>AI-specific security techniques protect N&S operations | Towards mission autonomy<br>AI continuously and reliably delivers on the business targets<br>Guaranteed AI functional safety | |
| Standards & Regulation | Lack of standards<br>Consultations with authorities and stakeholders | Emerging standards and basic interoperability<br>First compliant AI-based N&S automation solutions | Comprehensive standards and increased interoperability<br>Fully embedded policies and principles | |

# A multi-dimensional problem




* potentially multiple AI instances

# A multi-dimensional problem



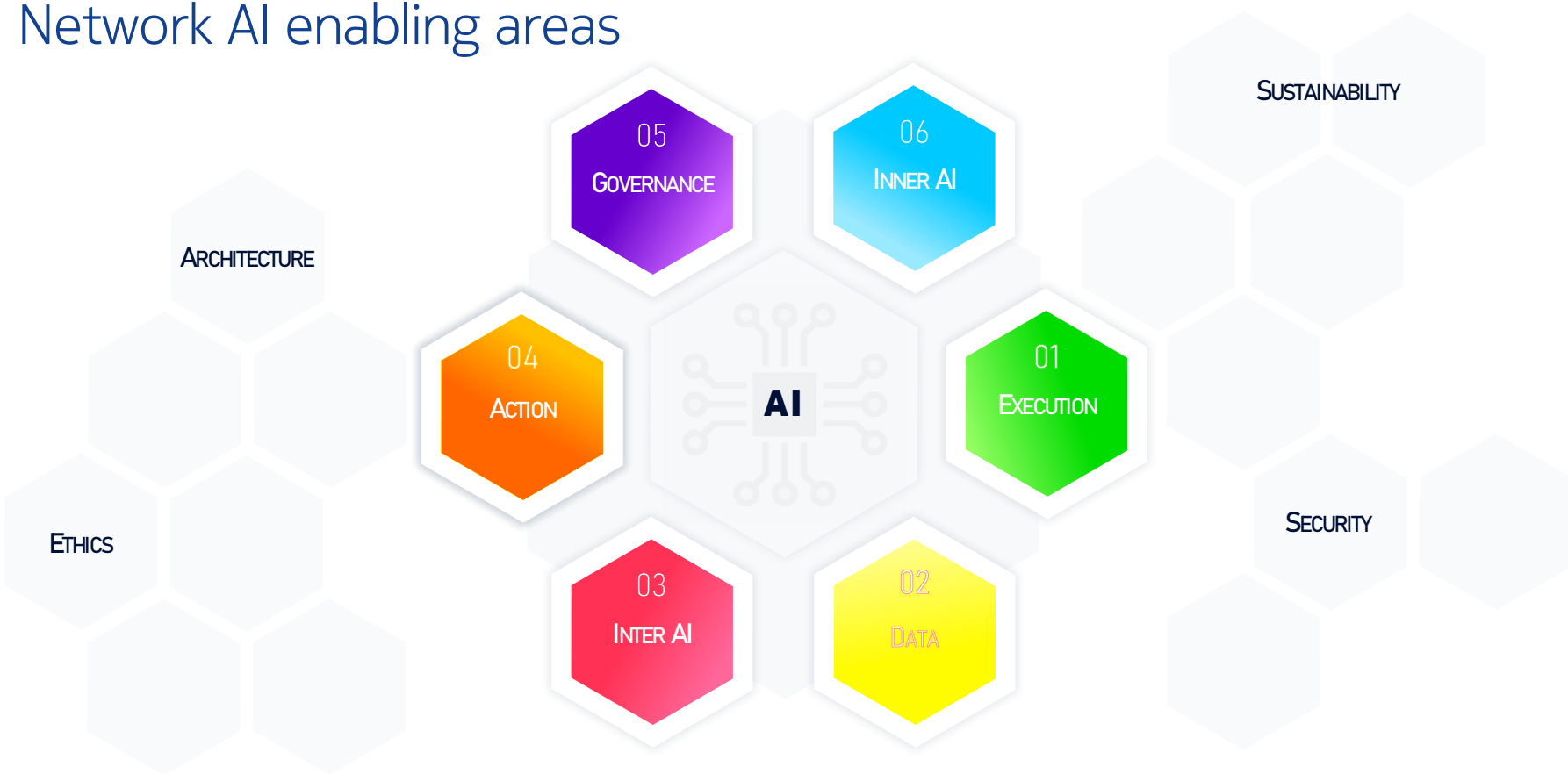
# Towards AI Native 6G

Generalize

Unify

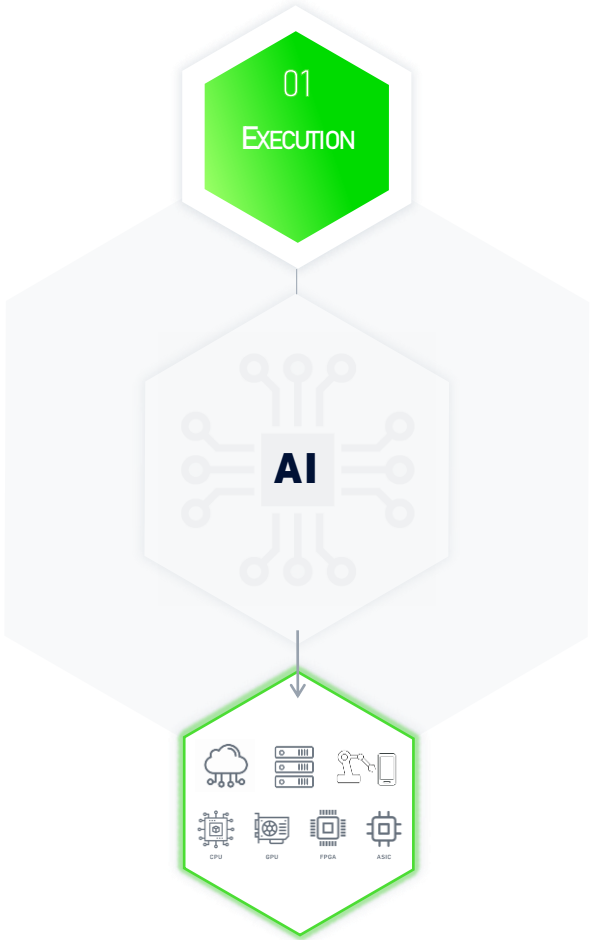
Expose

# Network AI enabling areas



## Considerations

- The infrastructure capabilities should support the AI application execution requirements and constraints (e.g. flexibility/efficiency) in any environment (e.g. on-cloud, on-premise, on-device)
- The environment composes a computing continuum, need to take into account the diversity of execution environment
- Different deployment options of the AI components driven for example by real-time requirements need to be supported, and planned consistently across domains




## Standards needs

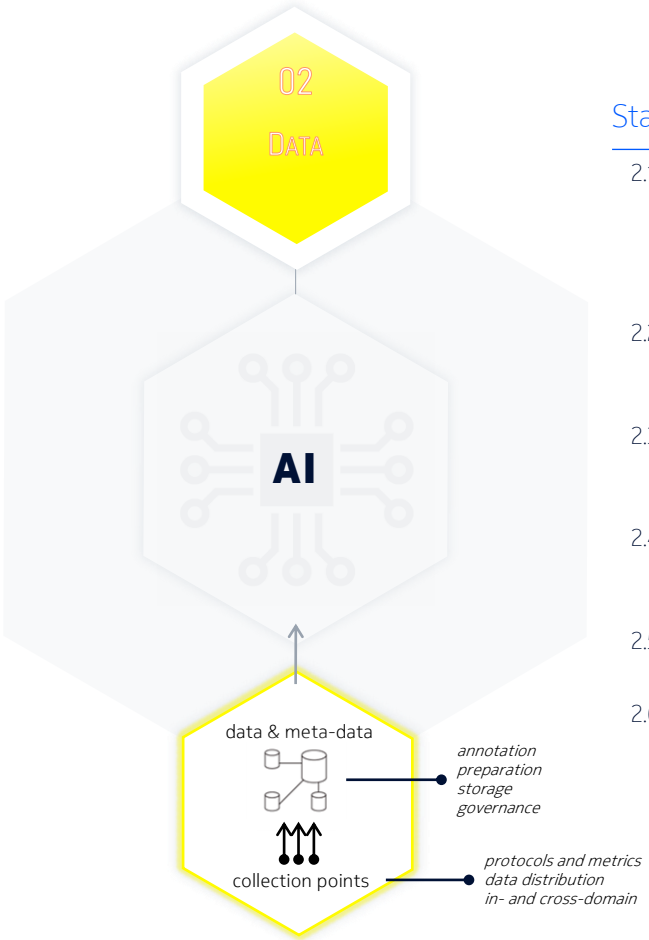
1.1 Ability to express requirements and constraints for deployment options of AI applications (affinity/placement/embedding) including distribution of components

1.2 Ability to dynamically adapt the provisioning and operations of AI components to take into account live network conditions, applications and solution requirements

1.3 Normalize the expression of AI application needs and capabilities, for both the training and inference phases

## Considerations

- Ensure access to the right (training and inference) data, at the right place, and at the right time
- Automation challenge when faced with lack of proper data (synthetic vs. operational data, limited access to data…)
- Data workflow encompasses annotation (and its automation), preparation, storage, collection points provisioning and available metrics
- Data needs for richer description and expressiveness
- Data tsunami balanced with intelligent data collection, addressing liberal to conservative approaches for data collection
- Data patterns are dynamic and change over time: limited validity of the learned model, limited generalization of the inferred knowledge (techniques tailored to a too specific use case or even dataset)
- Data governance with proper support and exposure for security and privacy, public vs. private data, anonymization, data encryption, etc.




## Standards needs

2.1 Ability for the AI application to express what data is required, under which form(s), pre-processing, storage requirements (persistency, availability, latency), discoverability and understanding of meta-data

2.2 Ability to convey semantics at different levels, as well as purpose to enable AI workflow automation and machine-to-machine operations

2.3 Support for flexible use of real and synthetic / simulated datasets, according to AI application data requirements

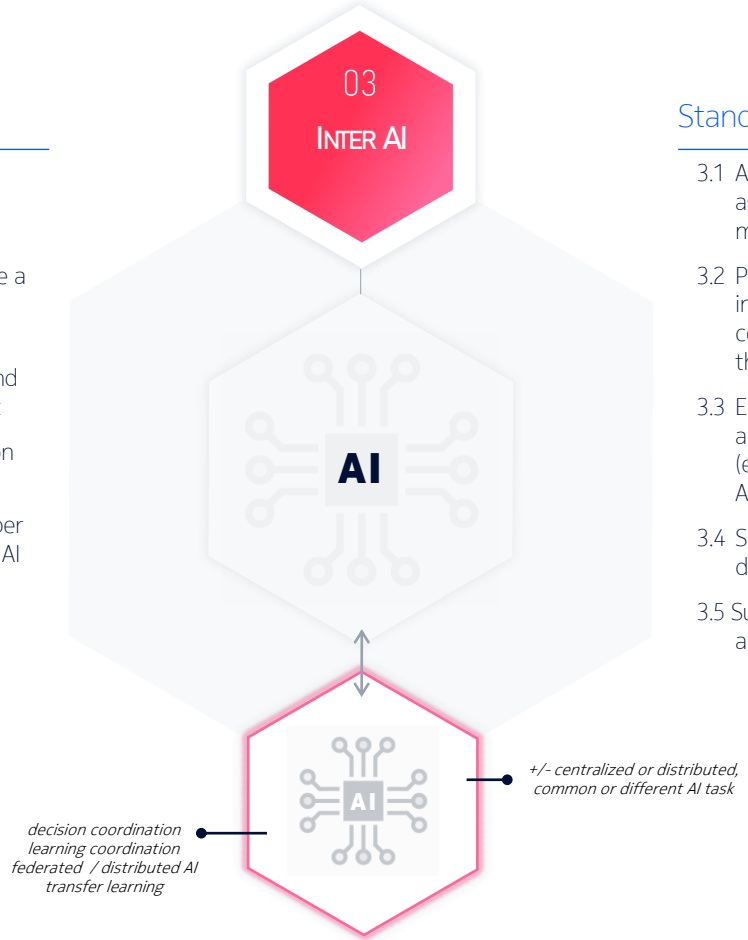
2.4 Augment data collection frameworks with automatically identified qualitative criteria and prediction capabilities

2.5 Support transfer of knowledge between domains, provide descriptions of domain characteristics

2.6 Utilize standards and best practices for data governance; extend as needed to support specificities of AI techniques together with security and privacy considerations

## Considerations

- Spectrum of learning approaches ranges from fully-centralized to fully-distributed
- AI applications can collaborate in learning different tasks or contribute collectively to solve a common problem
- Coordination between multiple, distributed AI applications is essential to ensure consistent and holistic operational view and means to act on it
- Applicability of transfer learning is dependent on the AI model semantics
- Trust relationship between AI applications, proper authentication validation and access control to AI applications operations




## Standards needs

3.1 Ability to convey semantics at different levels, as well as purpose to enable AI workflow automation and machine-to-machine (AI-to-AI) operations

3.2 Provide common logical interface and information/data models to enable inter-AI coordination and interaction, as well as sharing of the learning task description

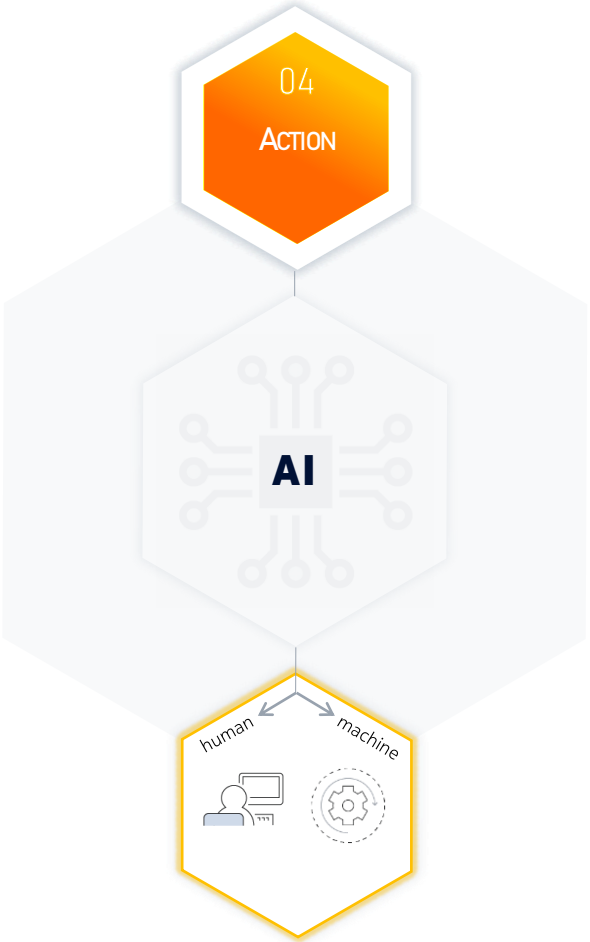
3.3 Ensure healthy balance in preserving business value and IP, and supporting an open innovation model (e.g. level of capability and details exposure to allow AI model swapping from one vendor to another)

3.4 Support transfer and re-use of knowledge, models descriptions and characteristics

3.5 Support building mutual trust between AI applications

## Considerations

- Case-tailored AI application and limited integration with the overall management activities pose challenges for end-to-end, AI empowered workflows and closed-loop automation deployments
- Currently, outputs of AI applications are not standardized
  - Output format and syntax depend on the AI model and implementation environment (e.g. different software libraries used)
  - Output semantics are AI applications specific
- AI application outputs can be descriptive, predictive and prescriptive; each with respective requirements for the decision-to-action automation (incl. human in the loop)
- AI applications outputs can target either human operators or machines
  - Outputs towards humans must be interpretable to allow decision enhancement and making output actionable.
- AI (ML) applications outputs are decoupled from domain knowledge, thus requiring intermediate step to contextualize, interpret and enrich AI outputs with other sources of information (e.g. human expertise, business intelligence…)




## Standards needs

4.1 Develop common description of AI application output, supporting:

a. Documentation: AI output can be automatically discovered or advertised, and linked appropriately in the management workflows ; AI output can also be used by human for design purposes

b. Identification: AI output can be automatically recognized and separated from any other output

Note: the description should be standardized, model- and semantic-based (e.g. ontologies) to enable per category AI model re-use ; AI output should contain proper meta-data describing e.g. accuracy

4.2 Enable support for different levels of action implementation such as informative (e.g. advice, suggestion, or recommendation), explicative (e.g. insight, hindsight, diagnosis) or imperative (e.g. prescription, command, authorization, or order)

4.3 Enable automatic, context-driven mapping between AI applications outputs and network/service/resource orchestration and control (closed-loop via actionable output)

4.4 Provide support in the transition phase involving human in the loop (e.g. for consent, verification, etc.)

## Considerations

### Usability

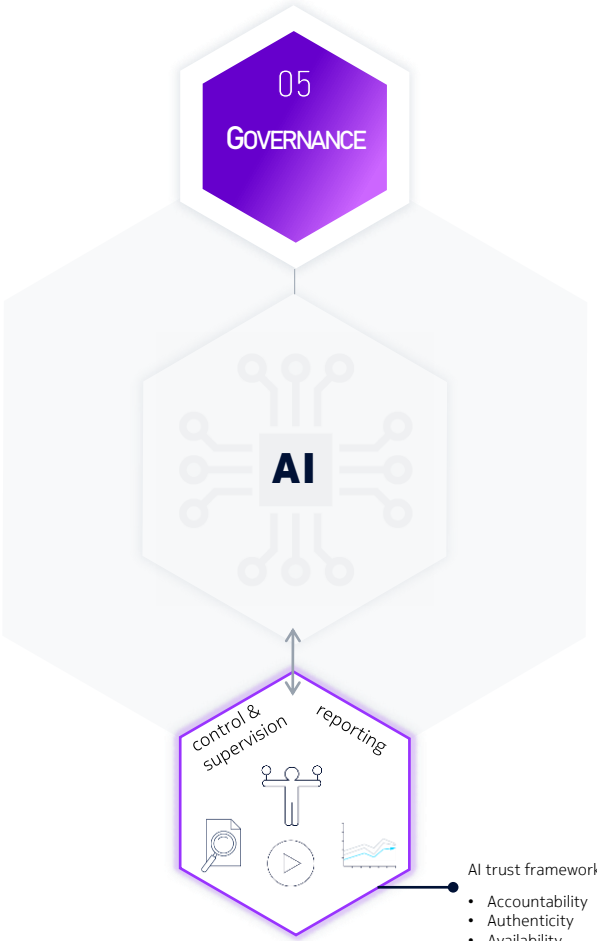
- AI should be easy to use, should reduce workload, and empower human operator with greater capabilities
- High-level, declarative goals (intents) hide complexity and simplify the setting of AI applications objectives, covering the comprehensive scope of governance aspects (security, privacy…) Current intent models, languages and functionality are not sufficiently developed and standardized. (e.g. for intent expression, translation, validation, assurance)
- Networking and AI expertise are different; new skills and tools are required for human operators to build, deploy, and tune AI applications during their entire lifecycle and automation process
- Different forms of human inputs must be supported (e.g. to define context for actions, to support unsupervised learning)

### Trust

- Networks are critical systems; AI results must be reliable, measurable, interpretable and accountable
- The AI applications outputs should be presented to the human operator in different forms adapted to the situation, to the level of trust and expertise, and to the needs of the human operator
- AI shall be able to provide information about its past, current and future activities: what has/is/will happen and the related reasons
- Human operator shall be able to verify and consent AI application decision

### Integration and operation

- AI applications should integrate and operate seamlessly within management functions, including appropriate and flexible composition and orchestration of AI applications functionality (e.g. data collection, knowledge extraction, inference, transfer…)




AI trust framework [1] ensuring:

- Accountability
- Authenticity
- Availability
- Integrity
- Privacy
- Quality
- Reliability
- Resilience
- Safety
- Security
- Transparency
- Usability

[1] according to ISO/IEC TR 24028:2020

## Standards needs

5.1 Specify a comprehensive and coherent set of intent-based capabilities to enable simple usage and support for full spectrum of governance functionalities

5.2 Support for feature-rich, Human-AI interface with dynamic Levels of Autonomy (LoA) and Levels of Supervision (LoS) enabling advanced control, supervision and reporting

5.3 Enable support for dynamic and adaptive Human-AI interactions providing guidance for human inputs, decision verification, appropriate information representation – to support the shift of human role with introduction of AI-empowered management

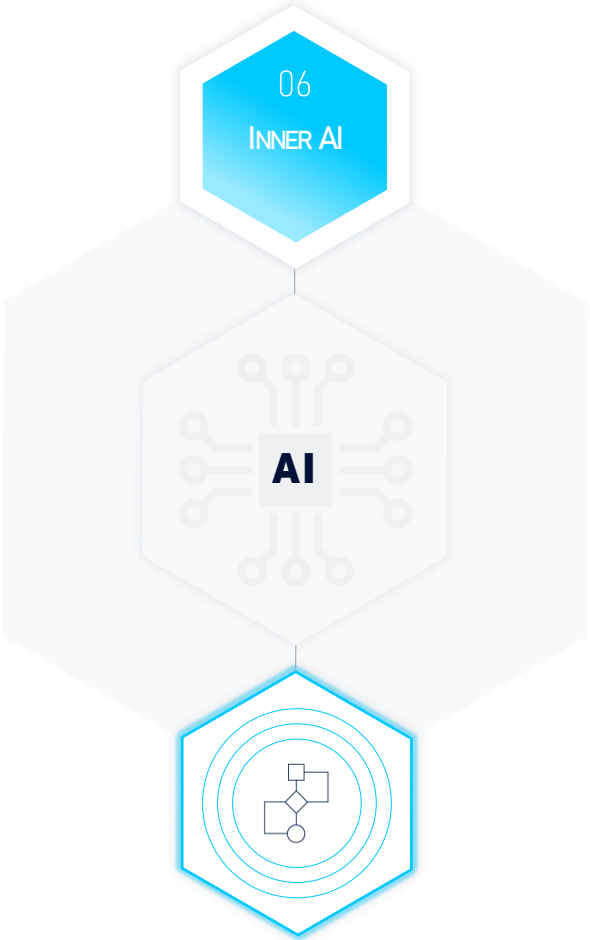
5.4 Support graceful fallback mechanism and procedures in case of AI "failure" e.g. by reverting to a more static and manual operation

5.5 Develop required "Trust as a Service" framework for full AI application features management, including performance, testing, validation, benchmarking, accountability, etc. based on self-measurement and self-evaluation

5.6 Develop N&S automation related AI applications lifecycle management, with corresponding management services and automation enablers for AI-centric build-deploy-run-terminate workflows; in-line with full Dev-(Net-Sec)-Ops cycle

## Considerations

- New types of algorithms or network-specific adaptation of algorithms may have new/different requirements and supporting needs from standards (driven by the evolution of AI technology)
- E.g. Machine Reasoning introduces novel requirements for 1) semantic data and knowledge representation, 2) algorithms integration with N&S automation frameworks, and 3) interactions with other forms of AI
- New use case categories (e.g. in the enterprise domain) may require adaptation of existing AI solutions (for CSP deployments) or development of new AI solutions
- AI-related aspects need to be modelled as part of a generic, multi-vendor closed loop-based N&S automation framework




## Standards needs

6.1 Provide flexible and extensible integration and operation capabilities to cope with requirements of new AI approaches

Note: as Inner AI is primarily focused on the algorithmic aspects, it is not in scope of standardization activities

# Recent advances in standards


**ETSI GS ZSM 012** [link]

Zero-Touch Network and Service Management
Enablers for AI-based Network and Service Automation

**O-RAN AIMLFW** [link]

O-RAN Alliance
AI / ML Framework

**NMRG AI Challenges** [link]

Network Management Research Group
Research Challenges in Coupling Artificial Intelligence and Network Management

# Summary

## Go Deep

**Maximize automation by integrating the full range of AI capabilities**

Research and specify key enabling, interoperable capabilities altogether at a level that allows composition and full lifecycle of AI-empowered N&S automation solutions

## Go Wide

**Support broadest application and deployment diversity across Telco + Verticals scenarios**

Ensure portability and reusability in various environments thanks to modular, extensible, open, service-based and model-driven APIs

## Go Safe

**Foster wide adoption and sustainable use of AI technologies for network automation**

Support transition of human role in control and supervision with means to interpret, evaluate and validate the behaviors of AI-based solutions

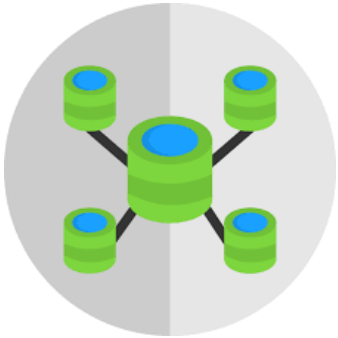
Develop governance and security frameworks to safeguard AI-driven operations and enable shift towards mission autonomy

# Future directions

## Sustainable AI

## Trustworthy AI

## Data at the core



# Call to action

- From use cases to more systematic approaches and generalization
- Methodology (reproducibility, reusability)
- New management/network architectures, models e.g. post NETCONF/YANG era
- AI for Network and Network for AI
- Data centricity, data quality
- More than performance. Think sustainability, reliability, scalability, trustworthiness, operability, **manageability**
- Tooling, data sets, AI challenges/playgrounds
- Community-wide and supported initiatives, focused areas
- More interdisciplinary research

# We are hiring AI Talents

https://nokia.ly/AITalent