

Information Involvement Influence

# AI - ETHICS, SECURITY, POLICY AND GOVERNANCE



**Chang Ge Director of CG Vision Consultancy** (Hon.) Senior Fellow, the University of Lincoln **Digital Transformation & Al Education** 

























The current generative phase

The aspirational/ emergent phase

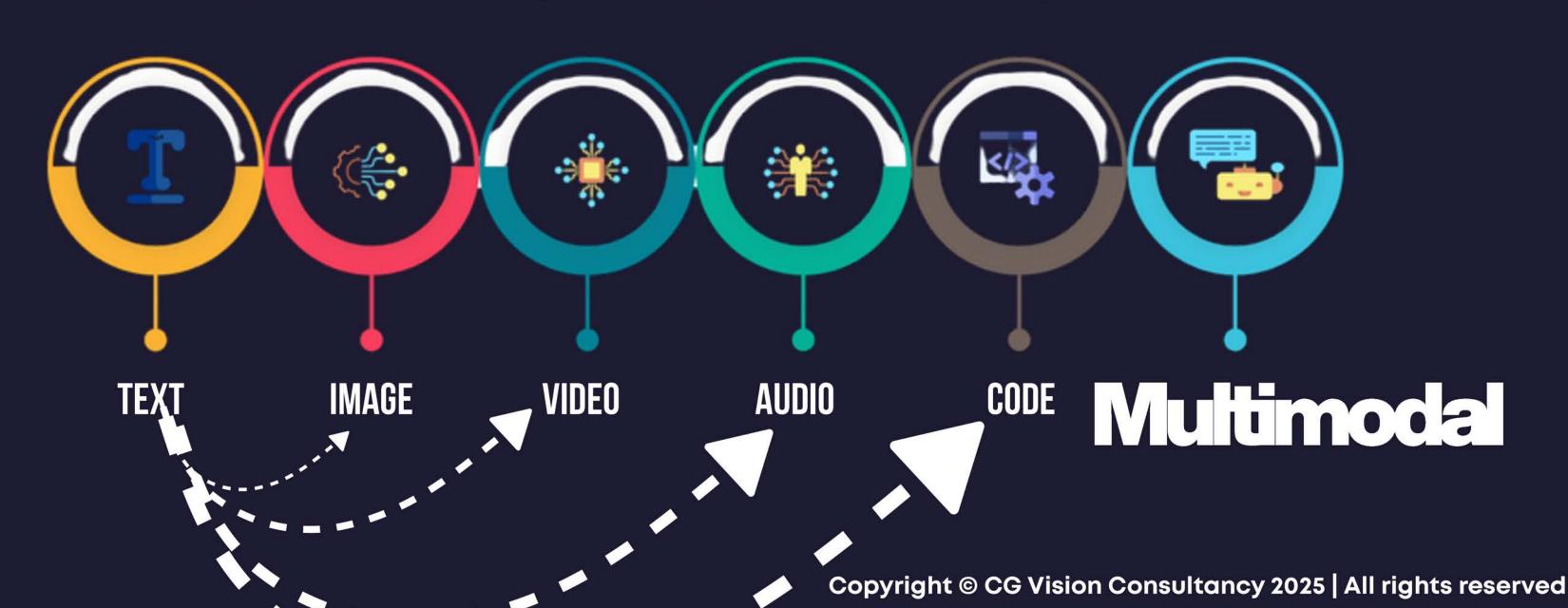
The integrative/ augmentative



AGI

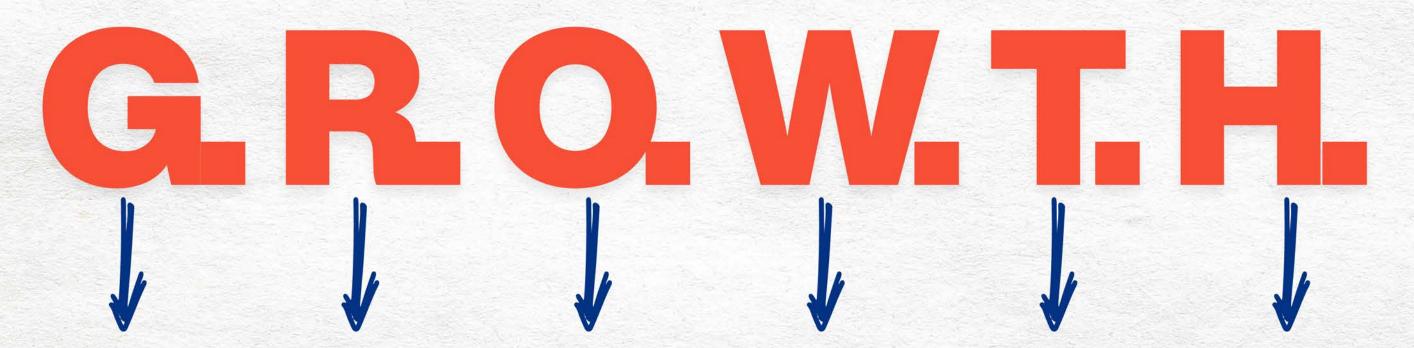








# Human Automation Agentic Al



Geo Revenue Operation Workforce Trust Holistic Sustantainble







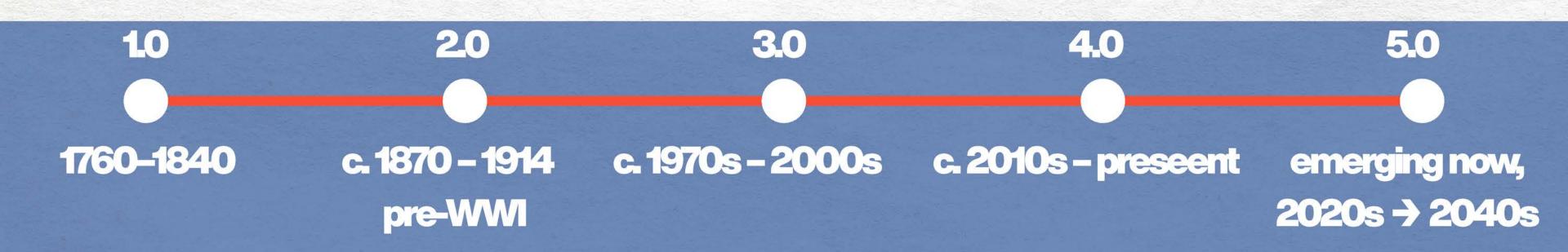


# Hope for Efficiency and Innovation

## Caution and Fear Around Autonomous Al



# Industry 5.0 Humaningenuity SocietyEnvironment









smart ovens

predictive maintenance



connected refrigeration

spare-parts demand forecasting

Gen Al in service documentation and customer journey

HACCP IoT sensors





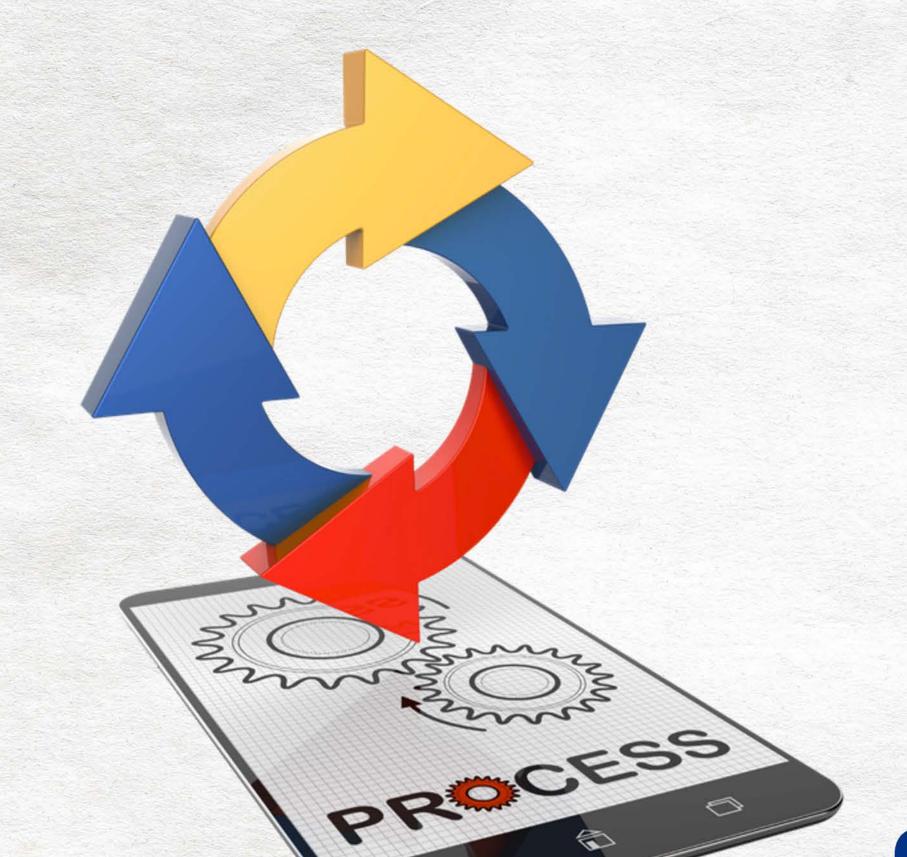
Safety

Privacy

Uptime

P

Brand Reputation



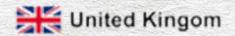






Adopt an Al Management System Fairness

Risk Management
Management
DPIA vs. GDPR



Legislation & regulation: Online Safety Bill (2022, draft); Data Protection and Digital Information Bill (2023, draft)

Standards: Algorithmic Transparency Standard (Central Digital Data Office,

Principles: A pro-innovation approach

to Al regulation (2023)

### CG VISION CONSULTANCY

#### Canada

Legislation & regulation: Directive on Automated Decision-Making (2019) Bill C-27: Digital Charter Implementation Act, including Al and

Data Act (AIDA) (2022, draft)

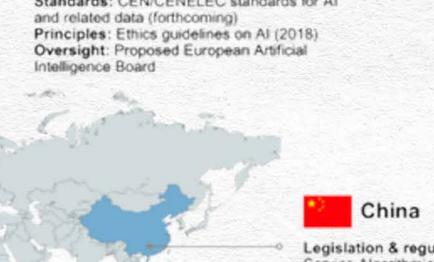
Standards: CAN-ASC-6.2: Accessible and Equitable Al Systems (2023, draft) Principles: Canada's Digital Charter

Oversight: Minister: Proposed Al and

Data Commissioner

#### **European Union**

Legislation & regulation: Proposed EU Al Act (2021, draft); Updates to the EU Product Liability Directive (2022, draft); Al Liability Directive (2022, draft); EU's Digital Services Act Standards: CEN/CENELEC standards for Al

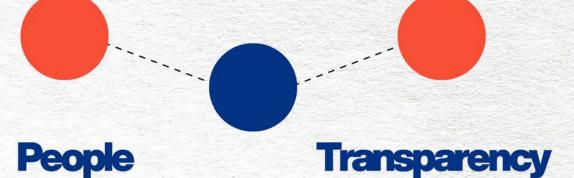


Legislation & regulation: Chinese Internet Information Service Algorithmic Recommendation Provisions (2021): Opinion on Strengthening the Ethics and Governance of Science and Technology (2022)

Standards: National Standards for Autonomous Vehicle Testing (2018)

Principles: New Generation Al Ethics Specifications (2019); New Generation Al Code of Ethics (2021); White Paper on Trustworthy Al (2021); Internet Information Service Algorithmic Recommendation Management Provisions (2021)





**Fairness** 

**Explainability** 

#### **United States**

Legislation & regulation: Federal Trade Commission Act, for deceptive practices from deepfakes or chatbots (1914);

Algorithmic Accountability Act (US AAA) (2022, draft)

Standards: NIST Al Risk Management Framework (2023)

Principles: Blueprint for an Al Bill of Rights

(2023

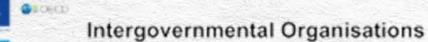


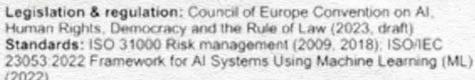
Legislation & regulation: Report and proposed substitute text for draft bills 5051/2019, 21/2020 and 872/2021 (2022, draft). Bill 705 on the compatibility of AI use in the public sector with ESG practices (2022, draft)

Standards: Incorporation of international standards National standards by the Brazilian Association of Technical Norms (ABNT)

Principles: Art. 3 of the proposed substitute text for draft bills 5051/2019, 21/2020 and 872/2021

(2022, draft)





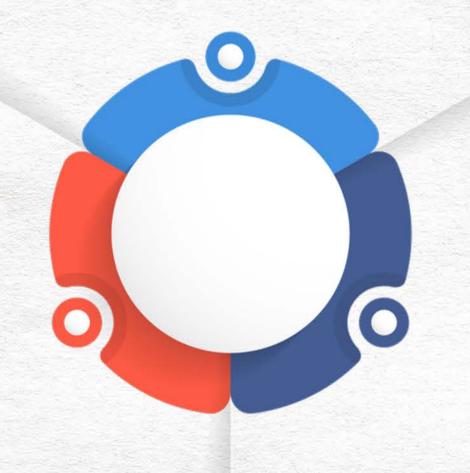
Principles: OECD Recommendation of the Council on Al (2019); UNESCO Recommendation on the Ethics of Al (2021)

(Examples of existing and emerging Al-specific regulatory approaches, **OECD, 2024)** 

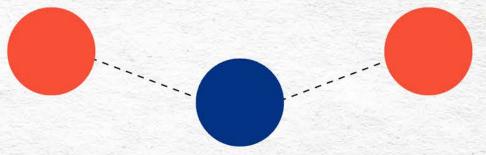


# Secure Al development lifecycle

# For connected equipment shipped into the UK







Secure by Design for Al and loT

Knowledge of cyber security & Al attack patterns



A sectoral approach by establishing cross-sectoral, non-binding principles.

**Regulate AI systems** 

across domains and

frameworks applicable to

applications with

comprehensive

all sectors.

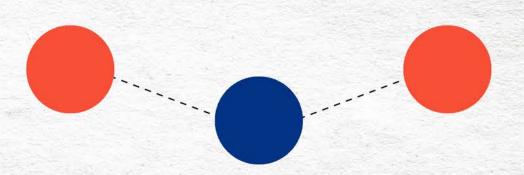
- UK Al Regulation: A Pro-Innovation Approach White Paper (2023)
- Data Protection and Digital Information (DPDI) Bill
- UK GDPR / Data Protection Act 2018
- Product Safety and Machinery Regulations (UK Supply of Machinery (Safety) Regulations)
- Equality Act 2010
- Al Safety Initiatives (Al Safety Summit, Al Assurance Frameworks by CDEI)

#### EU

- EU Artificial Intelligence Act (Al Act)
- EU Machinery Regulation (2023/1230)
- General Data Protection Regulation (GDPR)
- Digital Services Act (DSA) 2022
- EU Data Act (2022/2023)
- EU General Product Safety Regulation (GPSR) 2023
- Proposed AI Liability Directive
- Proposed Product Liability Directive (Revised)
- EU Ethics Guidelines for Trustworthy AI (2019)
- ISO/IEC 42001:2023 Al Management System Standard



# Policy & Regulation



What actually applies







Manage Risks are prioritized and acted upon based on a projected impact



## Governance

#### Level 1

The **SUM Values** are a set of ethical values intended to provide general normative guideposts and moral motivations for thinking through the social and ethical aspects of Al projects, and to establish well-defined criteria to evaluate their potential impacts as well as their ethical permissibility.

#### Level 2

At the second, and more concrete level, the SSAFE-D Principles are a set of actionable goals which provide tools to make sure that AI projects are bias-mitigating, nondiscriminatory, and fair, and that they safeguard public trust in their capacity to deliver safe, sustainable, accountable, transparent, and reliable Al innovation.

#### Level 3

At the third and most concrete level, the Process-Based Governance (PBG) Framework provides a mechanism for integrating the SSAFE-D Principles within Al design, development, and deployment processes as well as a template for documenting corresponding governance actions.

Adopt an Al Management System

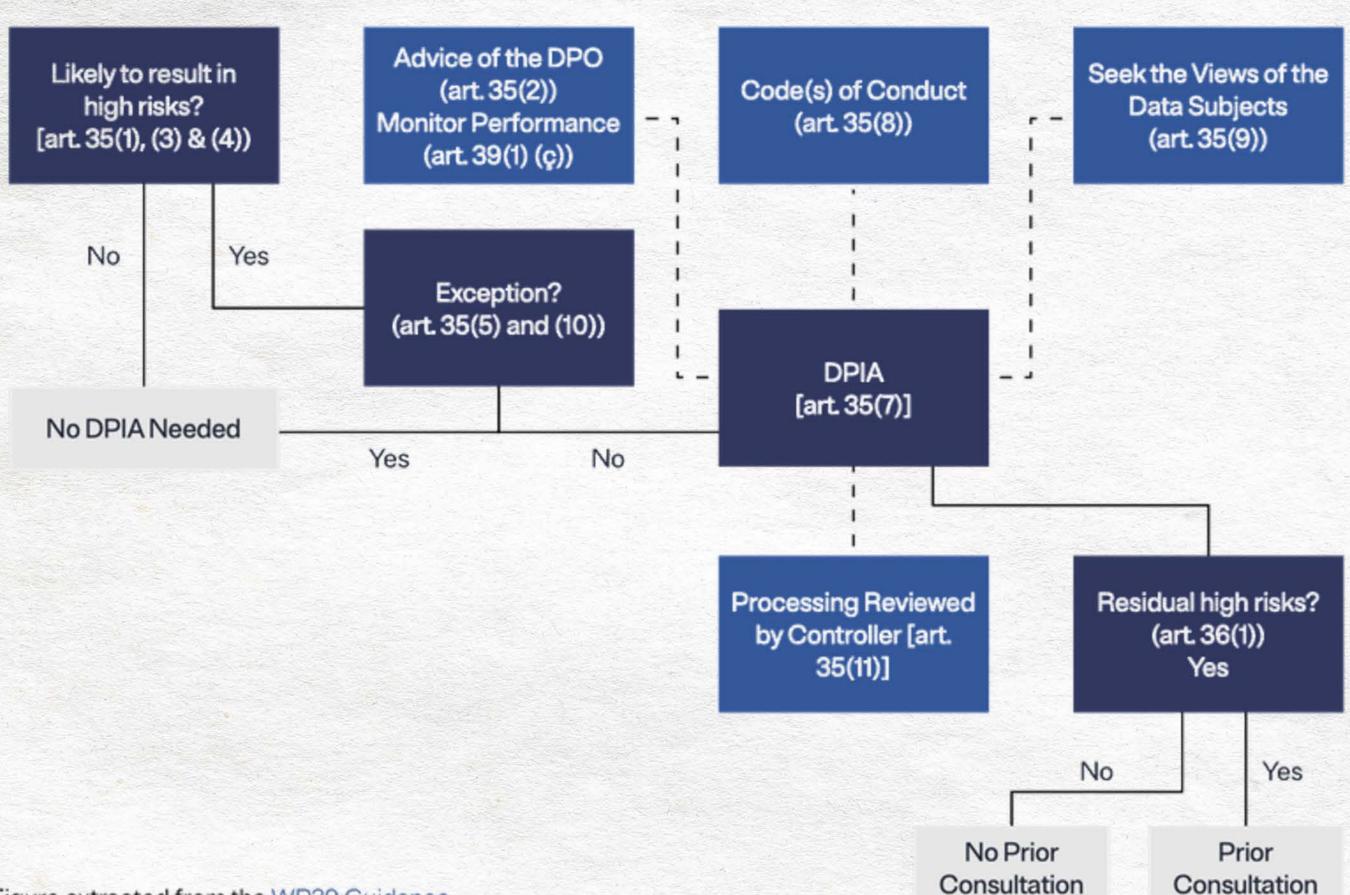
ISO420001

DPIA vs. Risk Management

Risk management The following figure illustrates the basic principles related to the DPIA in the GDPR:



CONSULTANCY



# Basic Principles related to DPIA in GDPR

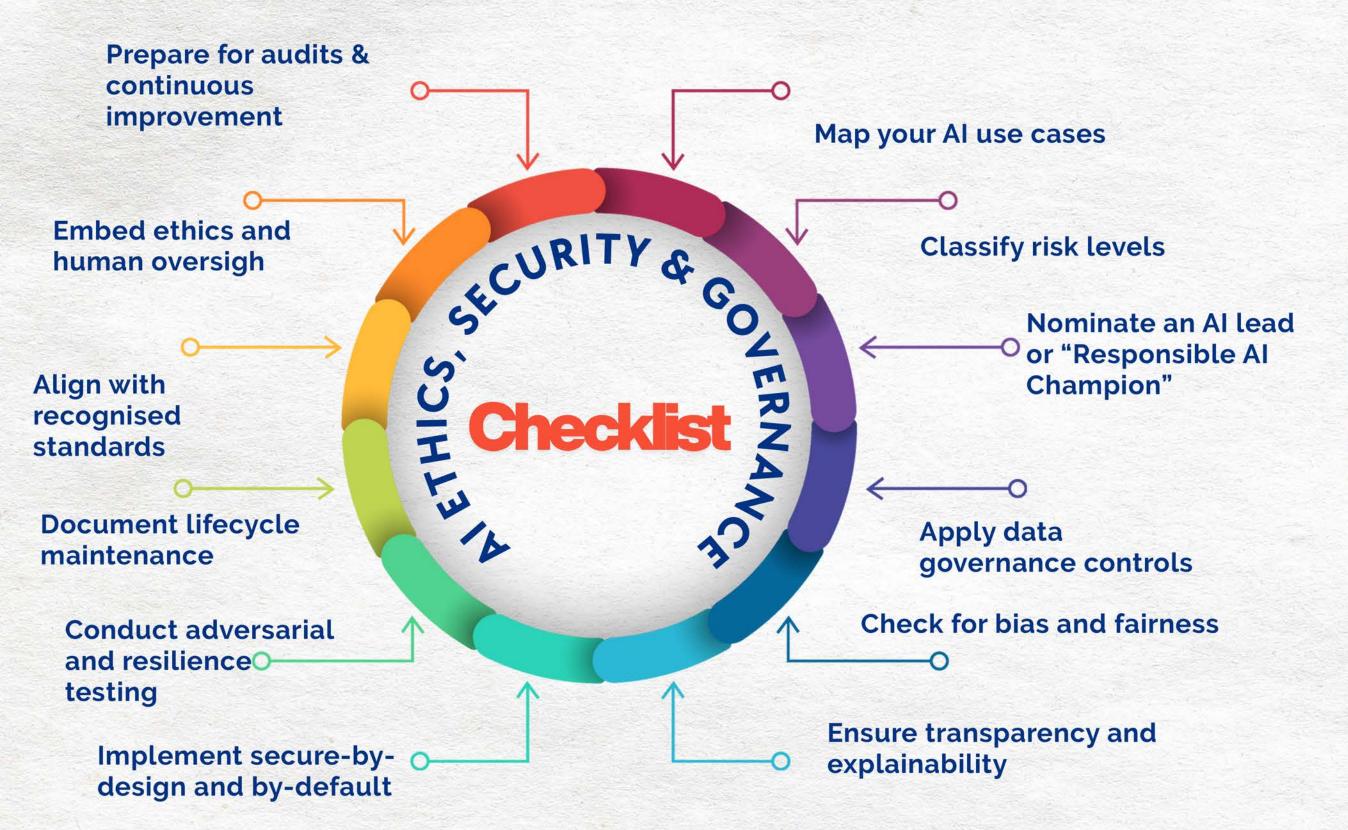


# Where's your biggest gap: Explainability, Cyber Security, Data Protection, or Documentation?



Start to commit one action before end of this week.









# THINK BIG | START SMALL BUILD | IMPROVE | GROW



scan here for feedback and AI adoption opportunities













