THINK for BL

Advancing public sector AI : It's time to move beyond talk and take action.

AI 공공분야

우리가 더 잘하려면?

더 이상 고고한 담론을 나누는데
낭비할 시간은 없다.

Park Ji Hwan | 박지환

- (Current) CEO of ThinkForBL Co., Ltd.
- (Current) Vice President, Korea Industry Intelligentization Association
- (Current) Chairman of the Digital Innovation Technology Committee, Korea Industry Intelligentization Association
- (Current) Board Member of TRAIN(Trustworthy AInternational Network) Korea
- (Current) Organizing Committee Team Leader of TRAIN(Trustworthy AInternational Network) Symposium
- (Current) Founder & Vice President of K.SEN (Korea Software Engineering Network)
- (Current) Korea Representative of ASQN(Asia Software Quality Network)
- (Current) Expert member of ISO/IEC SC 42 (Artificial Intelligence),
  the Ministry of Trade, Industry and Energy, Electrical and Electronic Information Division
- (Current) Director of the InnoBiz Association of the Ministry of SMEs and Startups
- (Previous) Program Committee Member of WCSQ(7th World Congress for Software Quality) Asia Oceania
- (Previous) Columnist for Information and Communication News, IT DAILY, Digital Chosun
- (Previous) Adjunct Professor, Department of Knowledge Information Engineering, Graduate School of Ajou University
- (Previous) Adjunct Professor, Department of Software Engineering, Jeonbuk National University

THINK for BL

R&D consulting for over 400 companies, performance of all national SW quality management projects. Unrivaled performance in patents, standards, and papers in the field of AI trustworthiness and education

2022~2024

Ministry of Land, Infrastructure and Transport, High Value-Added Convergence Logistics Delivery and Infrastructure Innovation Technology Development Project (KRW 2 billion)

TTA, Research Service for Deriving AI Trustworthiness Requirements (Public Society Field)

TTA, Research Service for Deriving AI Trustworthiness Requirements (Autonomous Driving Field)

TTA, Research Service for Deriving AI Trustworthiness Requirements (Medical Field)

TTA, Research Service for Deriving AI Trustworthiness Requirements (Super Large (Chat GPT))

TTA, Research Service for Deriving AI Trustworthiness Requirements (Security Field)

TTA, Research Service for Deriving AI Trustworthiness Requirements (Recruitment Field)

IITP, ICT Promotion and Innovation Foundation Creation Project (Research on AI Trustworthiness Education Plan for AI Developers))

2021

UIPA, Ulsan Genome Service Quality Consulting

Gwangju Information and Culture Industry Promotion Agency, AI Quality Improvement Consulting Support Service

NIPA, Vietnam ICT Market Entry Advisory Business Service Improvement Consulting

2020

NIPA, ICT Convergence Industry 4.0 (Shipbuilding and Marine) Project Quality Management Consulting

NIPA, GCS 6th Project B Company, Global Business Quality Consulting

2019

TTA, Development of a guide for localization of SW for export in Southeast Asia

NIPA, SW quality management for regional SW convergence product commercialization support project

UIPA, Marine Wiz platform infrastructure activation and hackathon planning consulting

2018

TTA, Development of a guide for localization of SW for export in the Middle East

NIPA, Quality management consulting for ICT convergence industry 4.0 (shipbuilding and marine) project

NIPA, SW quality management consulting for regional SW convergence product commercialization support project

2017

Daejeon Information and Culture Industry Promotion Agency, SW Convergence R&D Ecosystem Creation Project Quality Verification

NIPA, 2017 SW New Service Commercialization Strategy Establishment Consulting

TTA, Development of Export SW European Localization Guide

NIPA, ICT Convergence Industry 4.0(Shipbuilding&Marine) Project Quality Management Consulting

2016

NIPA, Quality Management Support and Consulting for Asset Redevelopment Projects

NIPA, SW Quality Management Consulting for Open SW Development Support Projects

NIPA, GCS School 'Code Quality Verification/Improvement Using Open SW Verification Tools'

KAI, Consulting for Establishing a Configuration Management System for the Development of Core Technologies for Light Armed Helicopters

KARI, Consulting for Establishing a DO-178C Requirements Traceability System Using Redmine

2015

NIPA, Quality Control and Test Support Consulting for Asset Redevelopment Projects

NIPA, Quality Control Consulting for SW Asset Bank Support Projects

TTA, Defense Acquisition Program Administration, Defense Acquisition Program Specialized Education

~2014

NIPA, Professional Company Development Type 1st Project Quality Inspection & Consulting

NIPA, Open SW Development Support Project Quality Management Consulting

NIPA, GCS Leading Technology Project Quality Management Consulting

NIPA, Professional Company Development and Exobrain Project Quality Management Consulting

NIPA, Field-Specialized Expert Mentoring Support Consulting for Securing Global Quality Capabilities in SW R&D Projects

NIPA, WBS 3rd Project Beta Test Consulting

NIPA, WBS 3rd Project SW Quality Management Consulting

NIPA, WBS 2nd Project Quality Management Consulting and Mentoring

Performed more than 100 other consulting

Registered over 80 patents

Received 8 Ministerial Awards

Established 8 standards

Intellectual Property Management Company (Recertification)

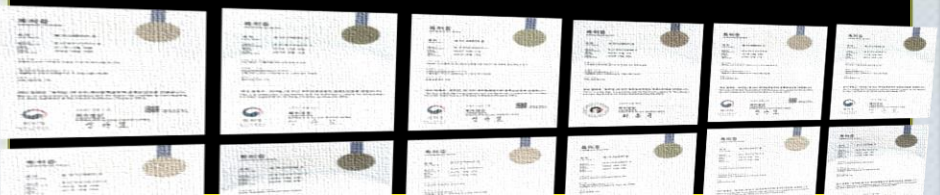
Excellent Invention Company (Recertification)

Companies with excellent employment & utilization of female R&D personnel

Family-Friendly Certified Company (Recertification)

Technology Innovation Small and Medium Enterprises (Inno-Biz)

Best-Value Service Company



The world is now thirsty for AI Trustworthiness.

세계는 지금, AI 신뢰성 확보에 목말라 있다.

Guidebook for Development of Trustworthy AI

하지만, 갑자기 왜?

But, why all of a sudden?

← ISO/IEC JTC 1

ISO/IEC JTC 1/SC 42

Artificial intelligence

About

Secretariat: ANSI

Committee Manager: Ms Heather Benko

Chairperson (until end 2024): Mr Wael William Diab

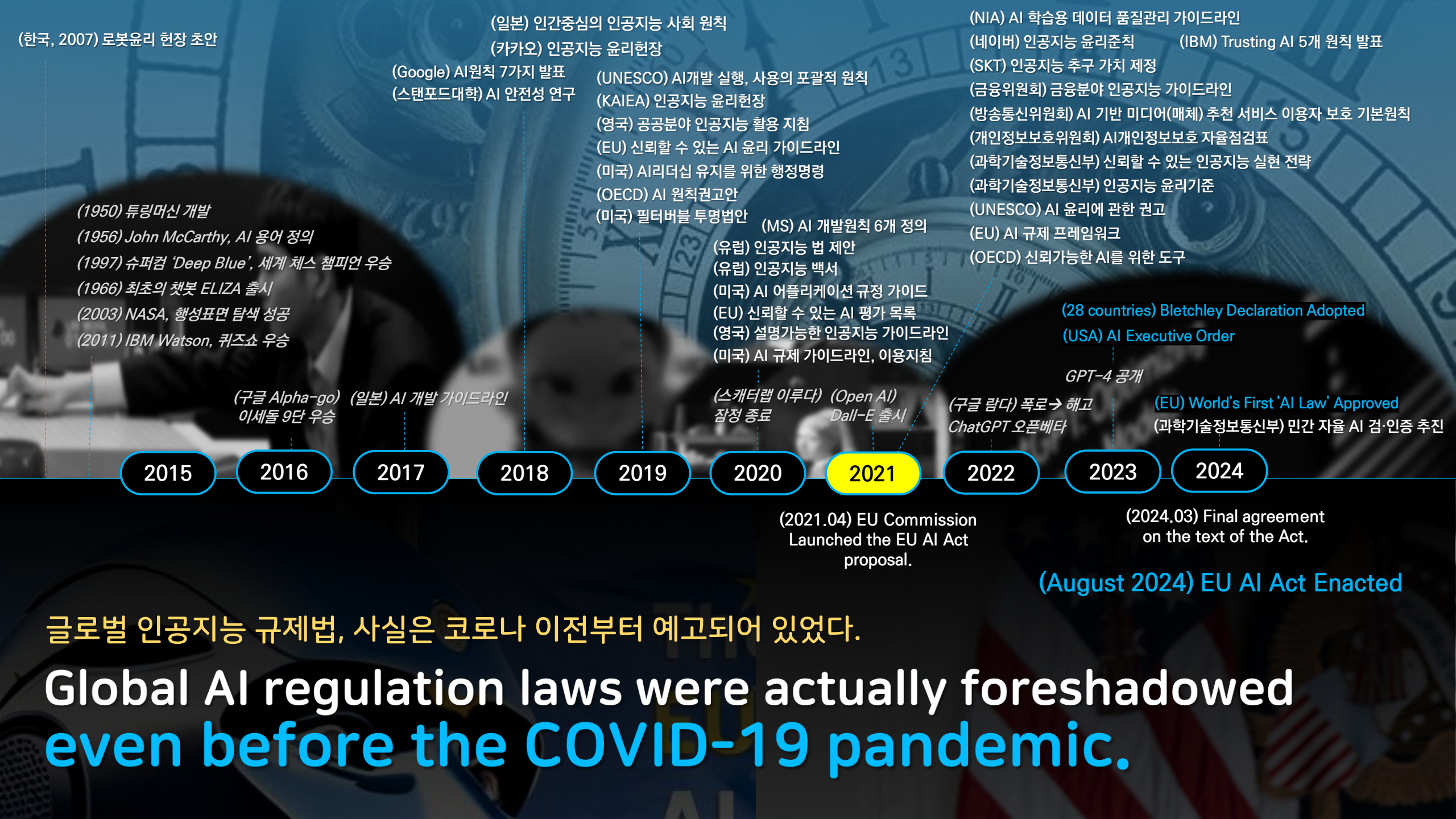
ISO Technical Programme Manager [TPM]: Mr Andrew Dryden
ISO Editorial Manager [EM]: Ms Jessica Navarria

Creation date: 2017

Scope

Standardization in the area of Artificial Intelligence

- Serve as the focus and proponent for JTC 1's standardization program on Artificial Intelligence
- Provide guidance to JTC 1, IEC, and ISO committees developing Artificial Intelligence applications



글로벌 인공지능 규제법, 사실은 코로나 이전부터 예고되어 있었다.

Global AI regulation laws were actually foreshadowed even before the COVID-19 pandemic.

What is the real purpose of AI regulation laws?

Between the lines of the regulation, their plans and ambitions are revealed.

AI 규제법의 진짜 목적은? 법안의 구석구석 행간에는 그들의 계획과 욕심이 있다.

▼ Provisions that could be interpreted as toxic clauses driven by national self-interest | 자국 이기주의로 인한 독소 조항으로 해석될 수 있는 조항

Article 51: Classification of General-Purpose AI Models as General-Purpose AI Models with Systemic Risk

Date of entry into force: According to: Inherited from:

제52조a 범용 AI 모델은 시스템적 위험이 있는 범용 AI 모델로 분류

SUMMARY +

1. A general-purpose AI model shall be classified as a general-purpose AI model with systemic risk if it meets any of the following conditions:

(a) it has high impact capabilities evaluated on the basis of appropriate technical tools and methodologies, including indicators and benchmarks;

(b) based on a decision of the Commission, ex officio or following a qualified alert from the scientific panel, it has capabilities or an impact equivalent to those set out in point (a) having regard to the criteria set out in Annex XIII.

2. A general-purpose AI model shall be presumed to have high impact capabilities pursuant to paragraph 1, point (a), when the cumulative amount of computation used for its training measured in floating point operations is greater than 10(^25).

Article 10: Data and Data Governance

Date of entry into force: According to:
2 August 2026 Article 113

▲이는 GDPR과 매우 밀접한 연관이 있으며,
EU의 데이터 보호 규정(GDPR)을 준수해야 하는 것은
EU 외부 기업들에게 큰 부담이 될 수 있음

1. High-risk AI systems which make use of techniques involving the training of AI models with data shall be developed on the basis of training, validation and testing data sets that meet the quality criteria referred to in

▲This is closely related to GDPR, and the requirement for companies outside the EU to comply with the EU's General Data Protection Regulation (GDPR) can be a significant burden.

2. Training, validation and testing data sets shall be subject to data governance and management practices appropriate for the intended purpose of the high-risk AI system. Those practices shall concern in particular:

이투데이

AI전략최고위협의회 가동…"EU AI법, 韓기업 '과징금 폭탄' 사전대응 시급"

Cumulative amount of compute used for its training measured in floating point operations (FLOPs) is greater than 10^25.

교육에 사용된 누적 계산량 부동 소수점 연산(FLOPs)에서 측정된 값은 10^25 (1 yottaflop) 보다 크다.

Threshold for GPAI with systemic risks : 10^{25}

Experts point out that the provisions of the EU's AI Act are still vague, but they advise preparation for the law since violations can result in fines of up to 7% of global revenue.

그 중에서도 특히 중요한 것은 바로 '책임'의 문제다.

Among them, the issue of 'responsibility' is particularly important.

In the era of AI, the concept of 'responsibility' changes.

Who will be held accountable for the mistakes of machines that make their own decisions?

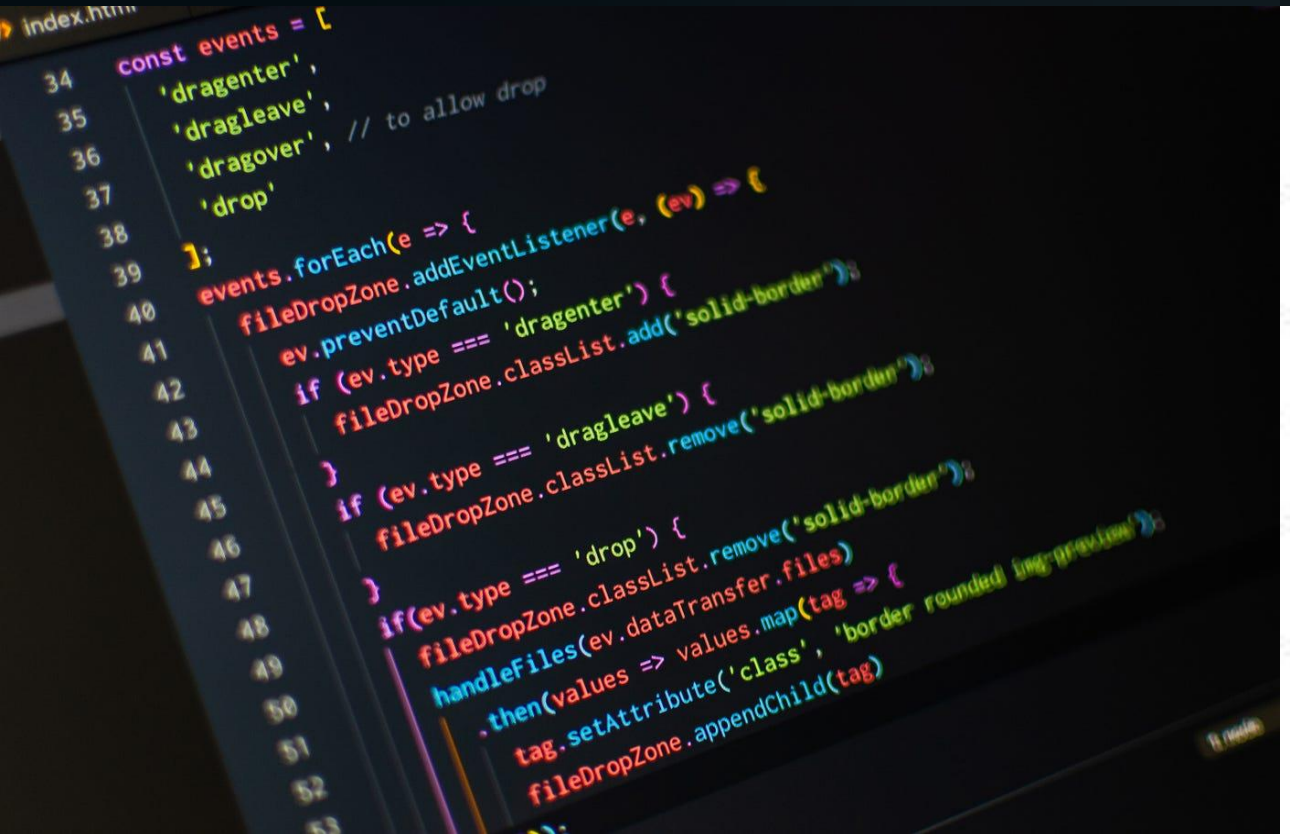
AI 의 시대에는 '책임'의 개념이 달라진다.
스스로 판단하는 기계의 잘못은, 누가 책임질 것인가?

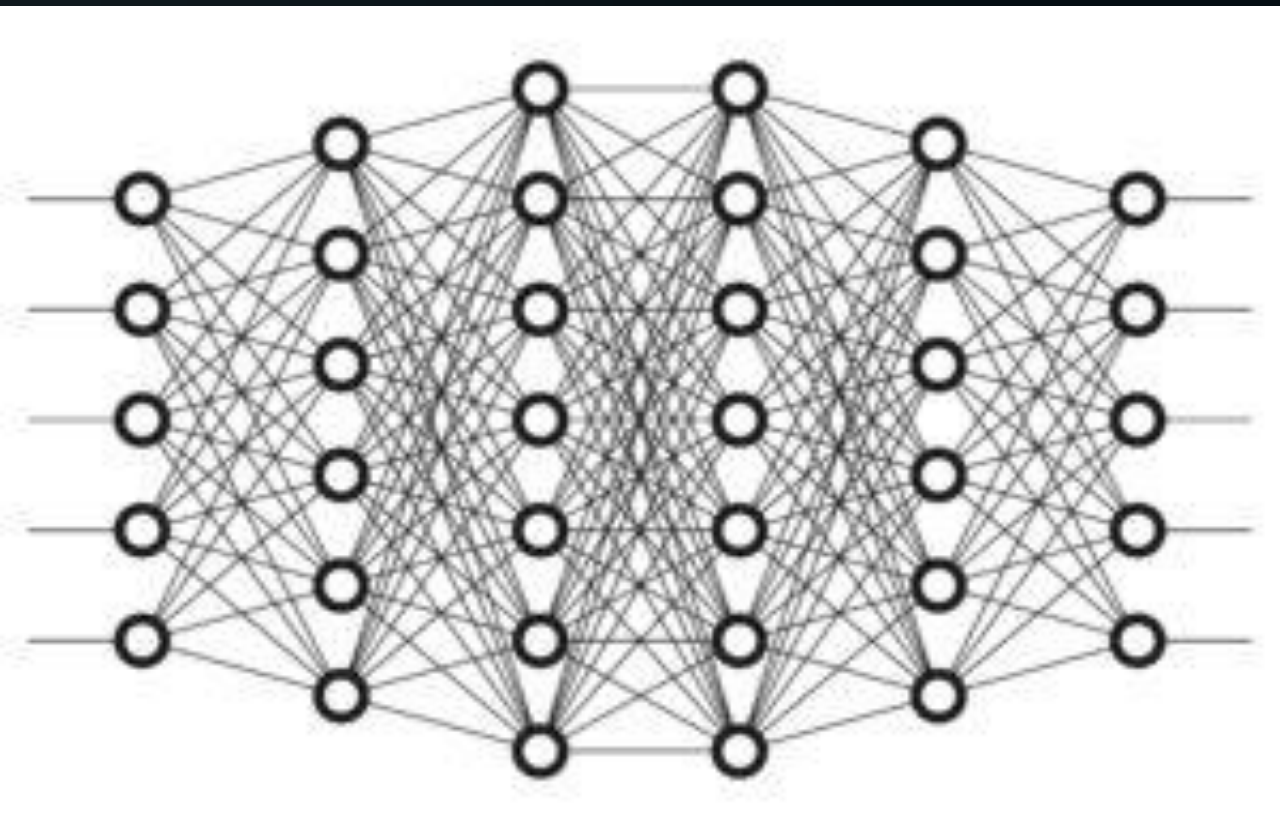
SW 1.0

Software operates according to the logic I have written.

SW 2.0

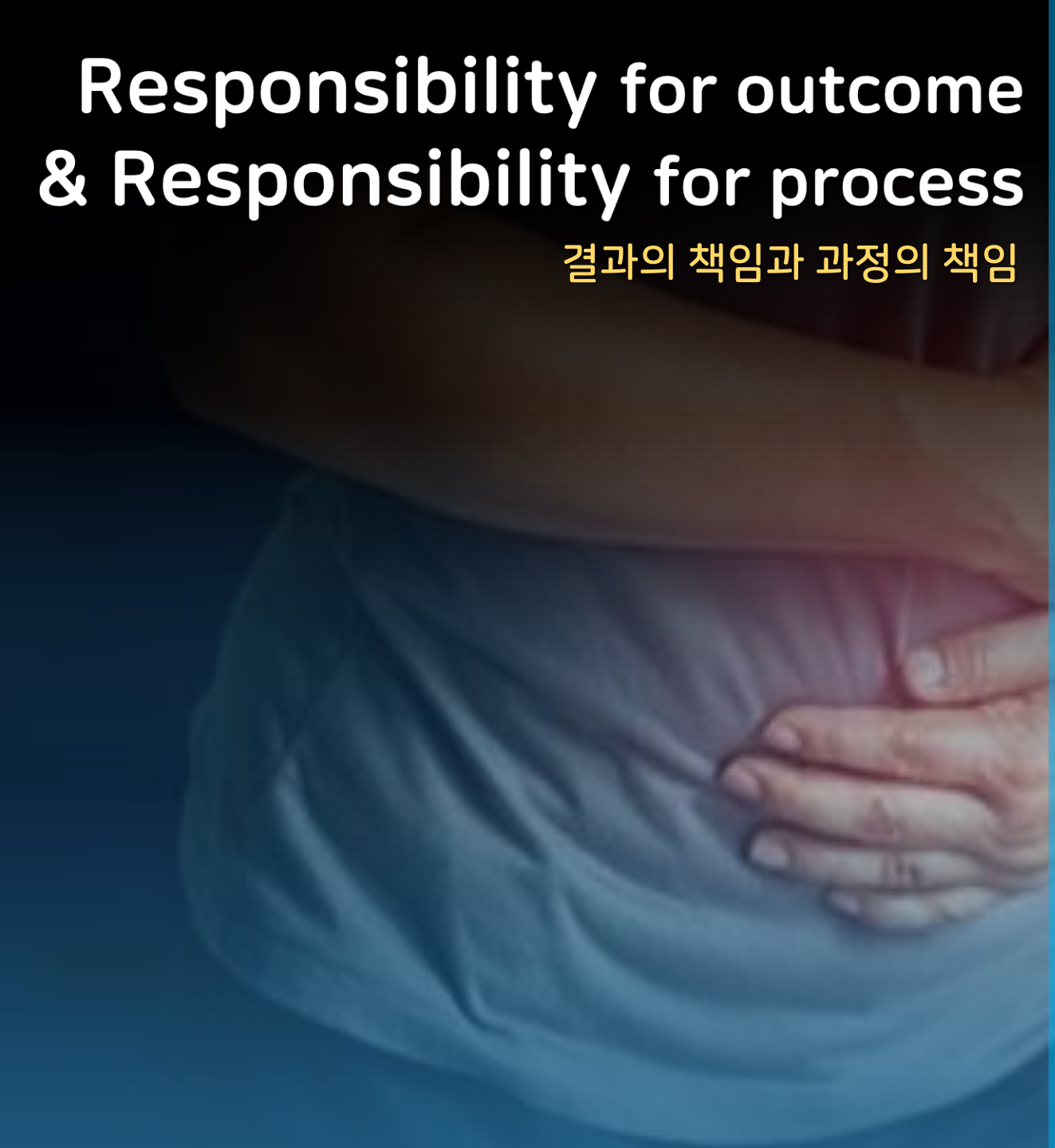
Software learns from data, develops its own criteria, and then makes decisions and acts on its own.





Responsibility for outcome & Responsibility for process

결과의 책임과 과정의 책임



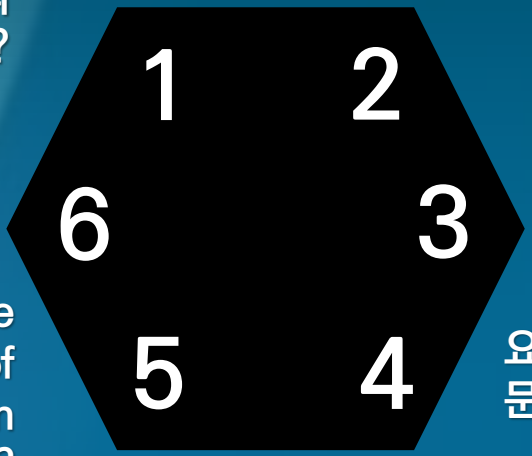
Is it an issue with the ingredients themselves?
식자재 자체의 문제인가?

식자재 유통 과정에서
변질되었기 때문인가?

Did the ingredients spoil during the distribution process?

음식을 먹은 사람의
부주의 때문?

Because of the carelessness of the person who ate the food?

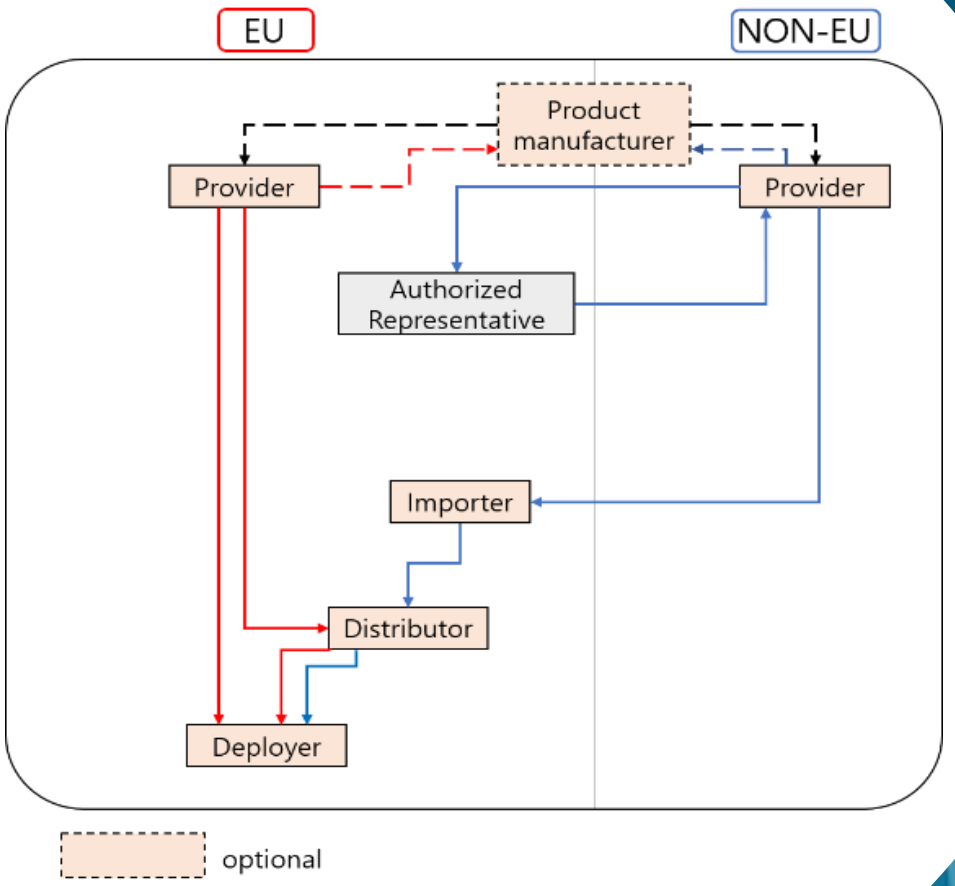


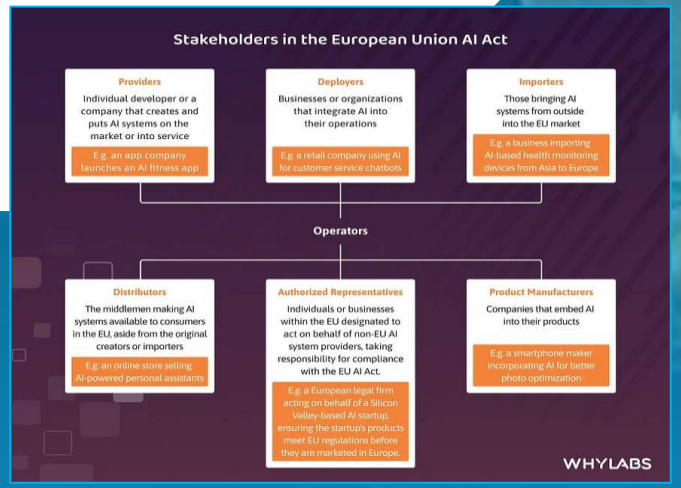
요리하는 과정에서
문제가 발생했는가?

Was there a problem during the cooking process?

음식을 제공하는 과정에서
문제가 생겼는가?

Did something go wrong in the process of serving the food?





Now, simply "delivering the specified product" is not enough to fulfill one's responsibility.

이제 '정해진 물건을 내놓았다'는 것만으로는 책임을 다할 수 없다.

Responsibility for process

Compliance with relevant regulations regarding hygiene, safety, and nutritional standards

위생•안전•영양상태 등 관련 필요 규정 준수

Responsibility for outcome

Completion of cooking within the designated time

정해진 시간 내 요리 완료

In the face of unpredictable upheavals, what will our choice be?

Will we move forward towards becoming an AI powerhouse?

예측이 어려운 격변 앞에서, 우리의 선택은?

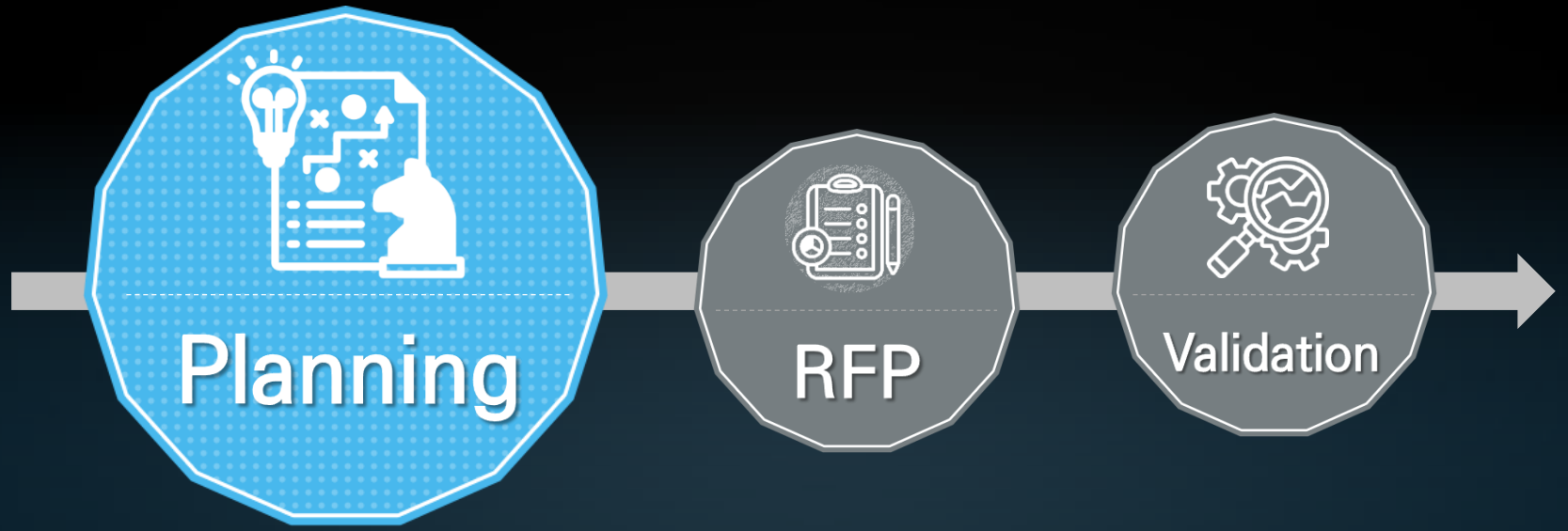
귀찮을 때는 폐지, 하지만 산업의 진보는?
AI 강국을 향해서 나아갈 것인가?

This is an opportunity to elevate our response level ▶
이것은 우리의 대응 수준을 한단계 높일 기회 ▶

In the era of AI, how to reduce the risk of legal disputes

Consider both foreign laws and future regulations

AI 의 시대, 법적 분쟁 위험을 줄이는 방법 | 해외의 법도, 미래의 법도 고려하라.



Review for any illegal elements or potential for disputes | 위법적 요소나 분쟁의 소지가 있는지 여부 검토

- ▶ Internet services should fundamentally be available without restrictions on nationality or residence.
- ▶ Excluding foreigners or overseas Koreans from administrative services could be seen as discriminatory.
- ▶ As all countries are competitively advancing AI legislation, there are limitations if only current legal standards are considered.
- ▶ A prioritization strategy is necessary: First, introduce services with minimal potential issues, and set service elements with legal risks as long-term goals.

- ▶ 인터넷 서비스는 기본적으로 국적이나 거주지 제한 없이 이용 가능해야 하기 때문
- ▶ 행정 서비스의 대상에서 외국인을 제외하거나, 재외 한국인을 배제한다는 것은 차별이라고 판단될 소지 존재
- ▶ 모든 나라들이 경쟁적으로 AI 법안 추진 중으로, 현재 법 기준만 고려해서는 한계가 존재
- ▶ 우선순위 전략 필요 : 문제 소지가 적은 서비스 선도입 후, 법적 리스크 있는 서비스 요소는 장기과제로 설정

In the Age of AI, How to Reduce Legal Dispute Risks

EU law requires you to keep your kitchen open.

AI 의 시대, 법적 분쟁 위험을 줄이는 법 | EU 의 법은 당신의 주방을 열어놓으라고 요구한다.

Reflecting Accountability Activities and Outcomes in RFP Requirements
| 책무성에 대한 활동과 결과를 RFP 에 반영하여 요구

Article 50 : Transparency Obligations for Providers and Deployers of Certain AI Systems

Article 11 : Technical Documentation

The technical documentation of a high-risk AI system shall be drawn up before that system is placed on the market or put into service and shall be kept up-to date.

Article 13: Transparency and Provision of Information to Deployers

High-risk AI systems shall be designed and developed in such a way as to ensure that their operation is sufficiently transparent to enable deployers to interpret a system's output and use it appropriately…

AI 검증 기술은 눈으로 봐서 알 수 없다

"AI validation technology cannot be understood by simply observing it."

Transparency Requirements under EU Standards

Planning
RFP
Validation

No domestic companies are currently equipped to handle this situation.

EU 기준이 요구하는 투명성
현 상황에서 이것을 감당할 수 있는 국내 기업은 없다.

ISO/IEC 12792 Six Categories of Stakeholders ← Information to be presented

AI Platform provider

AI 시스템이 배포되는 인프라
성능 최적화를 위한 모범 사례
지속 가능성 고려 사항

AI Service or Product Provider

AI 시스템 기본 사항
AI 시스템 요소
개발 및 배포되는 플랫폼의 제약 및 한계
시스템의 권장 사용 및 금지된 사용에 대한 정보
법률, 표준 및 인증 준수
응용 프로그래밍 인터페이스(API) 및 관련 문서

AI Producer

AI 시스템 설계 문서, 권장 필드 시험

AI developer

AI 모델 사용 요구 사항
성능 최적화를 위한 모범 사례
지속 가능성 고려 사항

AI Customer

운영 중 에너지 소비 및 에너지 효율성 향상 권장
설치 안내, 불만 처리 절차

AI Auditor

AI 시스템 품질 관리 시스템
특정 시스템 구성 요소에 대한 정보
관련 내부 요소에 대한 접근

AI System Level
AI Model Level
Dataset Level

이것을 모두 공개하라는 것은 '갑질'

Demanding full disclosure of all these levels is 'abuse of power'

How to Address Transparency Demands?

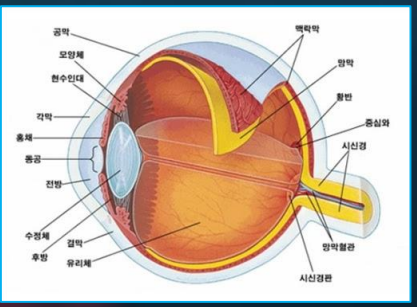
Insights from Belgian Experimental Research

투명성 요구에 대처하려면? | 벨기에의 실험 연구가 시사하는 것은

Planning

RFP

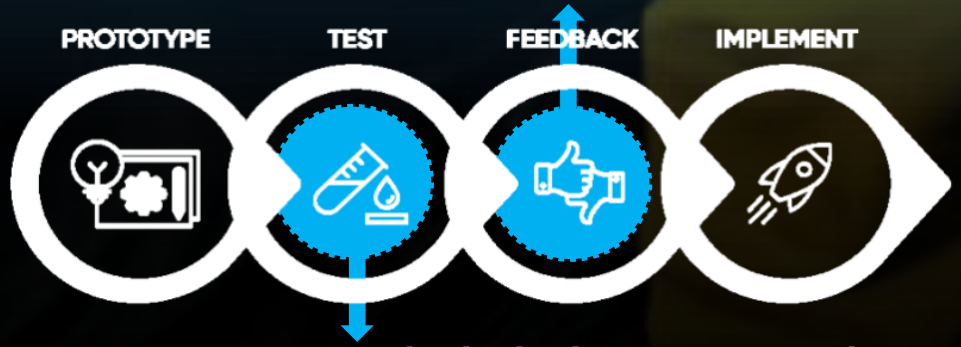
Validation



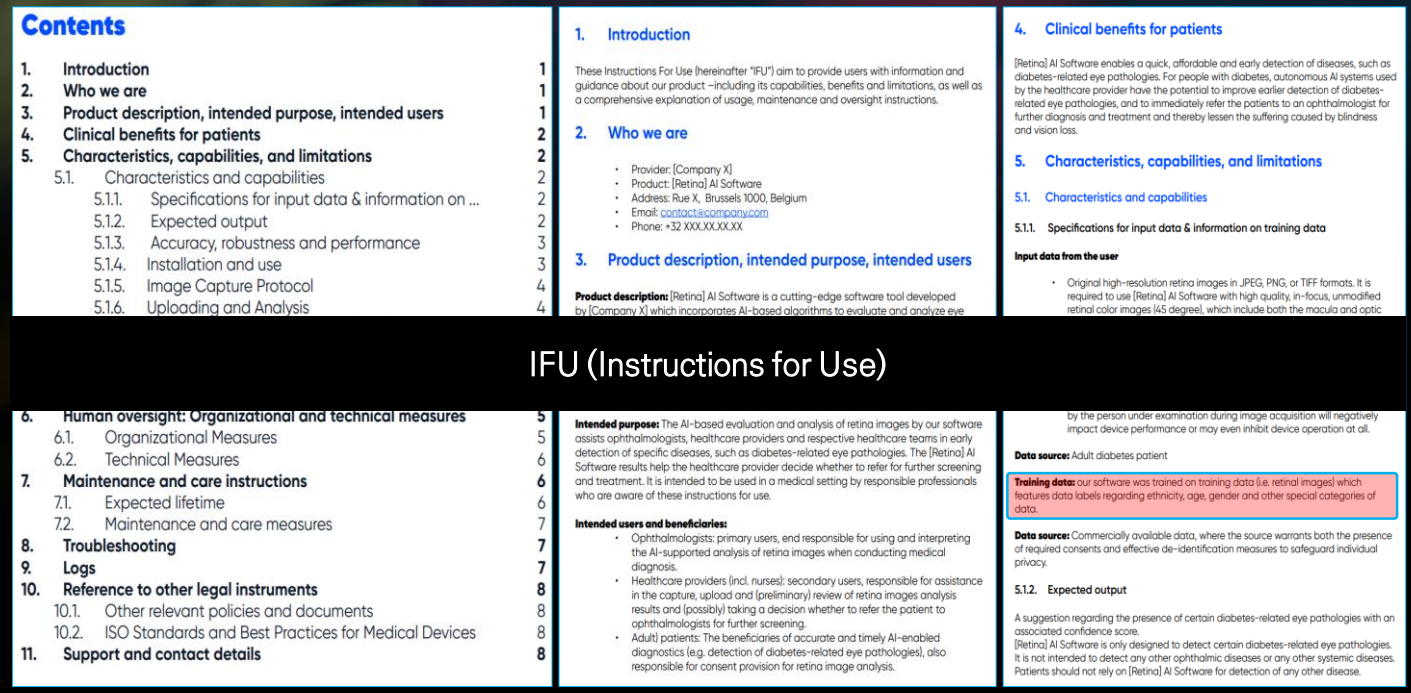
AI software for detecting eye pathologies based on retinal images

Total 15

- 2 AI developers/users representatives
- 4 consultants
- 4 legal experts
- 1 civil society organization representative
- 3 academic researchers with technical experience



- 17 Participants
- 4 Moderators
- 5 Groups



Contents

1. Introduction 1
2. Who we are 1
3. Product description, intended purpose, intended users 1
4. Clinical benefits for patients 2
5. Characteristics, capabilities, and limitations 2
   - 5.1. Characteristics and capabilities 2
     - 5.1.1. Specifications for input data & information on ... 2
     - 5.1.2. Expected output 2
     - 5.1.3. Accuracy, robustness and performance 3
     - 5.1.4. Installation and use 3
     - 5.1.5. Image Capture Protocol 4
     - 5.1.6. Uploading and Analysis 4
6. Human oversight: Organizational and technical measures 5
   - 6.1. Organizational Measures 5
   - 6.2. Technical Measures 6
7. Maintenance and care instructions 6
   - 7.1. Expected lifetime 6
   - 7.2. Maintenance and care measures 7
8. Troubleshooting 7
9. Logs 7
10. Reference to other legal instruments 8
    - 10.1. Other relevant policies and documents 8
    - 10.2. ISO Standards and Best Practices for Medical Devices 8
11. Support and contact details 8

1. Introduction

These Instructions For Use (hereinafter "IFU") aim to provide users with information and guidance about our product –including its capabilities, benefits and limitations, as well as a comprehensive explanation of usage, maintenance and oversight instructions.

2. Who we are

- Provider: [Company X]
- Product: [Retina] AI Software
- Address: Rue X, Brussels 1000, Belgium
- Email: contact@company.com
- Phone: +32 XXX.XX.XX.XX

3. Product description, intended purpose, intended users

Product description: [Retina] AI Software is a cutting-edge software tool developed by [Company X] which incorporates AI-based algorithms to evaluate and analyze eye

4. Clinical benefits for patients

[Retina] AI Software enables a quick, affordable and early detection of diseases, such as diabetes-related eye pathologies. For people with diabetes, autonomous AI systems used by the healthcare provider have the potential to improve earlier detection of diabetes-related eye pathologies, and to immediately refer the patients to an ophthalmologist for further diagnosis and treatment and thereby lessen the suffering caused by blindness and vision loss.

5. Characteristics, capabilities, and limitations

5.1. Characteristics and capabilities

5.1.1. Specifications for input data & information on training data

Input data from the user

- Original high-resolution retina images in JPEG, PNG, or TIFF formats. It is required to use [Retina] AI Software with high quality, in-focus, unmodified retinal color images (45 degree), which include both the macula and optic

IFU (Instructions for Use)

Intended purpose: The AI-based evaluation and analysis of retina images by our software assists ophthalmologists, healthcare providers and respective healthcare teams in early detection of specific diseases, such as diabetes-related eye pathologies. The [Retina] AI Software results help the healthcare provider decide whether to refer for further screening and treatment. It is intended to be used in a medical setting by responsible professionals who are aware of these instructions for use.

Intended users and beneficiaries:

- Ophthalmologists: primary users, end responsible for using and interpreting the AI-supported analysis of retina images when conducting medical diagnosis.
- Healthcare providers (incl. nurses): secondary users, responsible for assistance in the capture, upload and (preliminary) review of retina images analysis results and (possibly) taking a decision whether to refer the patient to ophthalmologists for further screening.
- Adult) patients: The beneficiaries of accurate and timely AI-enabled diagnostics (e.g. detection of diabetes-related eye pathologies), also responsible for consent provision for retina image analysis.

by the person under examination during image acquisition will negatively impact device performance or may even inhibit device operation at all.

Data source: Adult diabetes patient

Training data: our software was trained on training data (i.e. retinal images) which features data labels regarding ethnicity, age, gender and other special categories of data.

Data source: Commercially available data, where the source warrants both the presence of required consents and effective de-identification measures to safeguard individual privacy.

5.1.2. Expected output

A suggestion regarding the presence of certain diabetes-related eye pathologies with an associated confidence score.
[Retina] AI Software is only designed to detect certain diabetes-related eye pathologies. It is not intended to detect any other ophthalmic diseases or any other systemic diseases. Patients should not rely on [Retina] AI Software for detection of any other disease.

How to Address Transparency Demands?

Planning
RFP
Validation

It should start at the RFP stage.

투명성 요구에 대처하려면? | RFP 단계에서부터 시작해야 한다.

- User-Friendliness and Usefulness : 8 out of 9 reviewers rated positively.
- Transparency : 6 out of 9 acknowledged transparency in use cases.
- Structure & Accessibility : Improved accessibility with structured documents and focused paragraphs.
- Information Balance : Adequate information provided while avoiding overload.
- Clarity and Conciseness : Most respondents found the document clear and concise.
- Target Users : Clearly identified ophthalmologists as primary users.

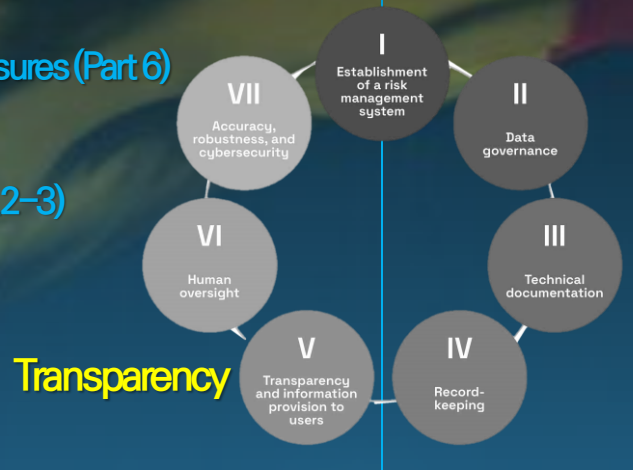
★ Distinguish between human supervision and technical measures (Part 6)

★ Specify retention period (Part 5.1.8)

★ Accuracy metrics and expected output information (Part 5.1.2–3)

★ See Troubleshooting section and additional manuals (Part 8)

- Technical Complexity of Terms : Terms like 'SaaS', 'broadband' are too technical.
- Unclear Main Users : Confusion between 'medical providers' and 'ophthalmologists'.
- Repetitive Expressions : Some expressions are redundant and confusing (Parts 3, 4).
- Lack of Input and Training Data Information : Insufficient details in section 5.1.1.
- Unclear Cybersecurity Measures : General description of cybersecurity measures, unclear responsibility allocation.

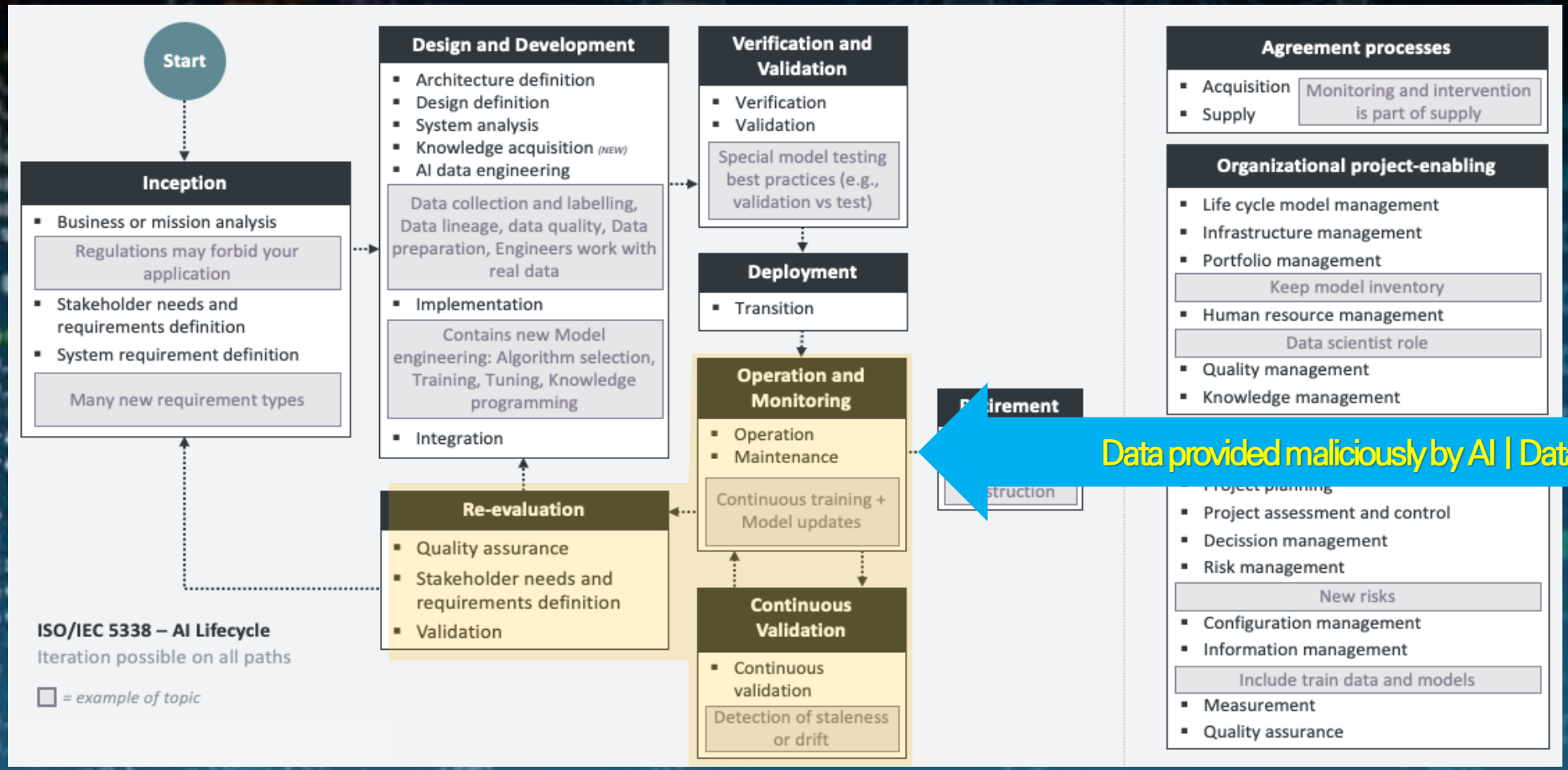


AI may cause new accidents even after release.

AI는 출시 이후에도 새로운 사고를 칠 수 있다.
SW 2.0 시대의 검수, 유지 보수가 달라진 점

Planning
RFP
Validation

Changes in Validation and Maintenance in the SW 2.0 Era



Data provided maliciously by AI | Data biased by social prejudice

Does having more data resolve bias?

Incorrect data can actually cause greater bias.

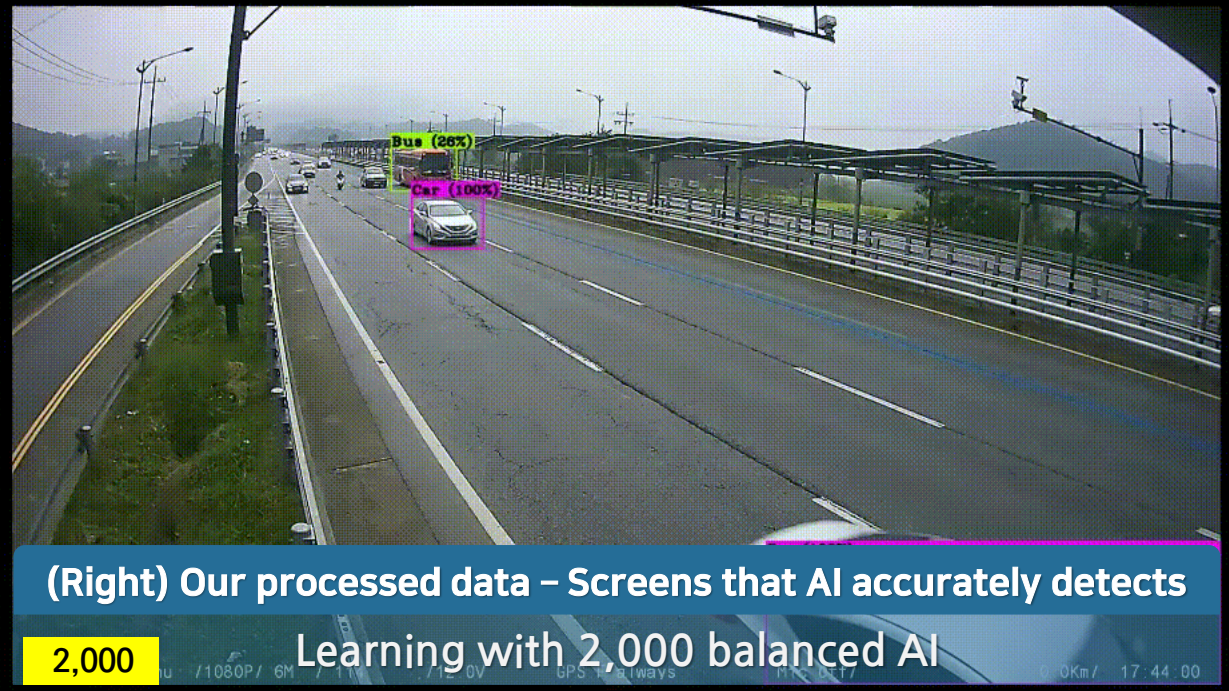
데이터가 많으면 편향이 해소된다? | 잘못된 데이터는 오히려 더 큰 편향을 일으킨다.

Planning

RFP

Validation

50,000



2,000

Load Truck

Trailer Truck

Bus

Mini Truck

Car

Original Dataset

Balanced Dataset

Raising Standards from the Public Sector

To win in the global market, you must exceed global standards.

공공 영역에서부터 기준을 더 높여야 한다. 세계 시장에서
승리하려면, 글로벌 기준 이상이어야 한다.

글로벌에서 기술력으로 인정받는 것이 '발전'이다.

Achieve global standards to be recognized for technological prowess.

Preconditions for Raising Standards

While we are criticizing Iruda for his slip of the tongue, they are preparing an elaborate plan.

기준을 높이기 위해 선행돼야 하는 것, 우리가 이루다의 말실수나 질타하는 사이에, 그들은 치밀한 계획을 준비하고 있다

Subdivide Safety Requirement Areas

Differentiate Application Areas by Service Nature and Target
For companies lacking capability, start with non-applicable areas.

안전성 요구 영역 세분화

To create technology that meets the highest level of demand

Establishment of institutional support plan

제도적 지원 방안 마련

Public & Private Markets

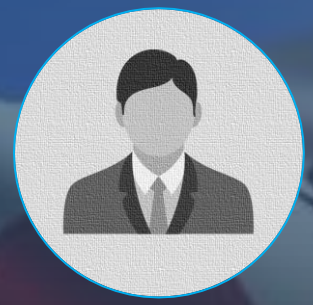
Innovations not possible in the private market where short-term results are crucial.

단기 성과가 중요한 민간 시장에서는 할 수 없는 혁신이 있다.

Adequate Budget

High-level requirement

Knowledge



Expert

Business

공공이 깃발을 들면, 기업은 달려간다. 협력해서 경쟁력을 키우지 않으면 국제 경쟁에서 이길 수 없다.

When the public sector leads, companies will follow.

To succeed in international competition, cooperation is essential to enhance competitiveness.