

The background features a dark blue gradient with a glowing horizontal line. On the left, there are golden circuit board patterns. On the right, there are blue network-like patterns with nodes and lines. The text 'REAL' is in large white letters, and 'SUMMIT 2023' is in smaller white letters below it.

REAL

SUMMIT 2023

SAMSUNG SDS

위험은 줄이고 기회는 최대한으로

금융 AI 활성화로 가는 안전장치, AI 거버넌스 사례와 전략

삼성SDS 금융 컨설턴트 조남용 프로

AGENDA

I. AI, 기회와 리스크

II. 쟁점 및 고려사항

III. 거버넌스 수립방안

I.

AI, 기회와 리스크

AI, 기회와 리스크를 바라보는 시각의 공존



기회에 대한 강조



Bill Gates

“ AI가 엄청난 편익을 가져다 줄 것이라는 점은 분명하다 ”

* Reuters Interview ('23.4.4)



Yann LeCun

“ 리스크도 발생하지만, 한편으로 막대한 가치를 창출한다고 생각한다 ”

* venturebeat.com ('23.4.7)



리스크에 대한 우려



Geoffrey Hinton

“ AI로 인해 실존적 위험에 직면하게 되는 시대가 왔다 ”

* MIT Technology Review ('23.6.12)



Yoshua Bengio

“ AI 제공자들이 하는 일이 무엇인지 정부가 나서서 추적, 감시할 수 있어야 한다 ”

* BBC News ('23.5.31)

리스크, 사전관리가 필요하다는 공감대 증가

AI 리스크 사례 | 금융 · 공공 · 서비스

The Washington Post ('19.1)

애플카드 알고리즘,
Goldman Sachs에 대한 성 편견 혐의

신용 평가에서 성차별,
금융소비자에 대한 설명 미흡

POLITICO ('22.3)

네덜란드 세무당국,
알고리즘 사용위험성을 유럽에 경고

인종, 출신에 대한 편견,
아동 기본권 침해

The New York Times ('23.5)

변호사가 ChatGPT를
사용할 때 일어난 일

Hallucination,
사회적 비용 증가, 책임소재 공방

AI 리스크 사례 | Foundation Model

Foundation Model 제공자의 EU AI Act 준수 수준 측정 ('23.6)





Draft AI Act Requirements	GPT-4	Cohere Command	Stable diffusion v2	Claude	PaLM 2	BLOOM	LLaMA	Jurassic-2	Luminous	GPT-NeoX	Totals
Data sources	● ○ ○ ○	● ● ● ○	● ● ● ●	○ ○ ○ ○	● ● ● ○	● ● ● ●	● ● ● ●	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	22
Data governance	● ● ○ ○	● ● ● ○	● ● ○ ○	○ ○ ○ ○	● ● ● ○	● ● ● ●	● ● ● ○	○ ○ ○ ○	○ ○ ○ ○	● ● ● ○	19
Copyrighted data	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	7
Compute	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	● ● ● ●	○ ○ ○ ○	● ○ ○ ○	● ● ● ●	17
Energy	○ ○ ○ ○	● ○ ○ ○	● ● ● ○	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	● ● ● ●	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	16
Capabilities & limitations	● ● ● ●	● ● ● ○	● ● ● ●	● ○ ○ ○	● ● ● ●	● ● ● ●	● ● ● ○	● ● ● ○	● ○ ○ ○	● ● ● ○	27
Risks & mitigations	● ○ ○ ○	● ● ○ ○	● ○ ○ ○	● ○ ○ ○	● ● ● ●	● ● ● ○	● ○ ○ ○	● ● ○ ○	○ ○ ○ ○	● ○ ○ ○	16
Evaluations	● ● ● ●	● ● ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ● ○	● ● ● ○	● ● ● ○	○ ○ ○ ○	● ○ ○ ○	● ○ ○ ○	15
Testing	● ● ● ○	● ● ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ● ○	● ● ● ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	10
Machine-generated content	● ● ● ○	● ● ● ○	○ ○ ○ ○	● ● ● ●	● ● ● ○	● ● ● ○	○ ○ ○ ○	● ● ● ○	● ○ ○ ○	● ● ● ○	21
Member states	● ● ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	● ● ● ●	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ○ ○ ○	○ ○ ○ ○	9
Downstream documentation	● ● ● ○	● ● ● ●	● ● ● ●	○ ○ ○ ○	● ● ● ○	● ● ● ○	● ● ● ○	○ ○ ○ ○	○ ○ ○ ○	● ● ● ○	24
Totals	25/48	23/48	22/48	7/48	27/48	36/48	21/48	8/48	5/48	29/48	

* Source : 스탠퍼드대학교 Foundation Model 연구센터

II.

쟁점 및 고려사항

해외 쟁점 | EU AI Act (브뤼셀 효과)

리스크 단계		의무 사항
Unacceptable 	<ul style="list-style-type: none"> • 신체/정신적 위해, 감정/실시간 생체인식 • 자연인에 대한 점수화, 평가, 분류 	사용금지
High 	<ul style="list-style-type: none"> • 보건/안전/환경에 위험이 될 수 있는 부품/제품/서비스 • 채용/인사/직업훈련 등 기본권에 영향 	위험관리 시스템 구축, 감독당국에 문서 보고, 추적 기록 등 투명화, EU 적합성 CE마크 부착
Limited 	<ul style="list-style-type: none"> • 인간과 상호작용, 생체 데이터 기반 감정 예측 • 콘텐츠를 생성하거나 조작 	AI와 상호작용 및 콘텐츠의 AI 생성 사실 공개
Minimal 	<ul style="list-style-type: none"> • 기타 그 밖의 AI 서비스 	자율적 행동강령 수립 및 준수
Foundation Model	<ul style="list-style-type: none"> • 대규모 데이터로 사전 학습된 모델 • 다른 AI 개발의 기초 모델 	거버넌스 시스템 구축, 불법 콘텐츠 생성 방지, 학습 데이터 출처 명시, 편향성 검증

국내 쟁점 | 정책 · 법안 (활성화 & 신뢰 확보)

현황 >>> >>>

시사점*



정책

인공지능의 활용 촉진 · 부작용 최소화

- AI R&D 전략 ('18.5) → AI 국가전략 ('19.12)
- AI 법·제도·규제 정비 로드맵 ('20.12)
- 부처별 AI 가이드라인 발표 및 후속 조치 추진 중



법안

AI 산업육성 및 신뢰기반 조성에 관한 법안

- 7개 발의된 유관 법안 통합, 상위법 성격
- > 선허용/후규제, 고위험 AI 정의, 사업자 책임



소비자

소비자 피해예방 · 권리보장

- 정보주체의 권리 명시 및 피해구제 방안
- 분야별 · 영역별 구체적인 고위험 인공지능 분류

>>> >>>

> Two Track Approach 필요

- 정책수립, 규제개선 기초에 근거한 **혁신 추진**
- 산업별 주무 부처의 **가이드라인에 따른 자율 규제**

> 규제의 불확실성 대비

- AI 서비스 · 상품 **유관 현행 법률과 연계성을 우선고려**
- > 금융소비자 보호법, 신용정보법, 개인정보보호법

> 소비자 보호에 선제적 준비

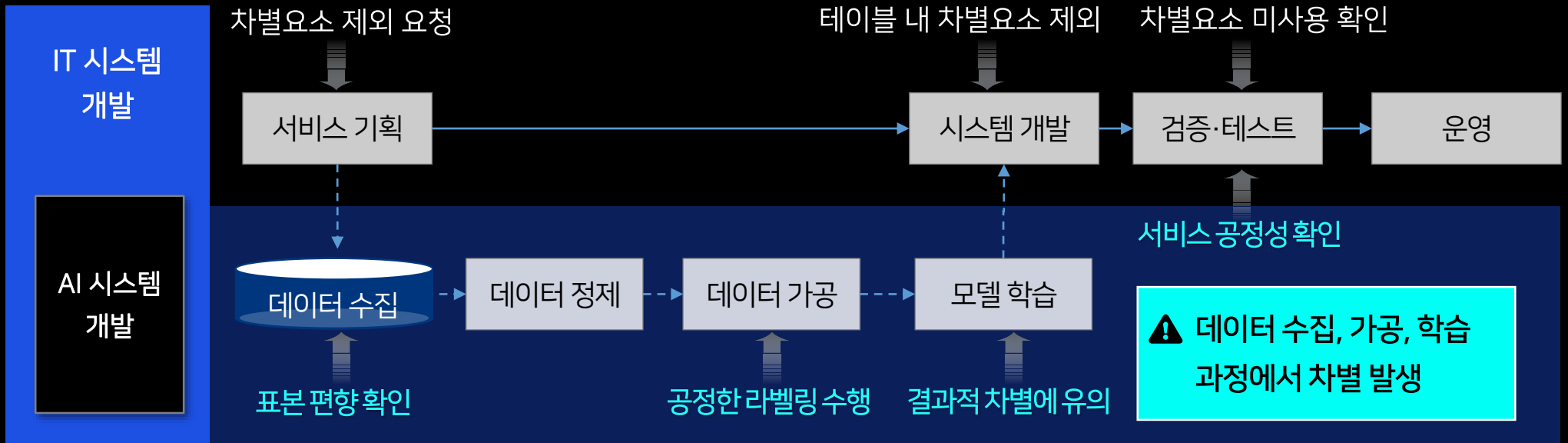
- 업 특성을 고려한 **소비자 이해, 권리 존중**

* ref. www.samsungsds.com/kr/insights/ai_governance.html

고려사항 | ① 소비자 차별금지

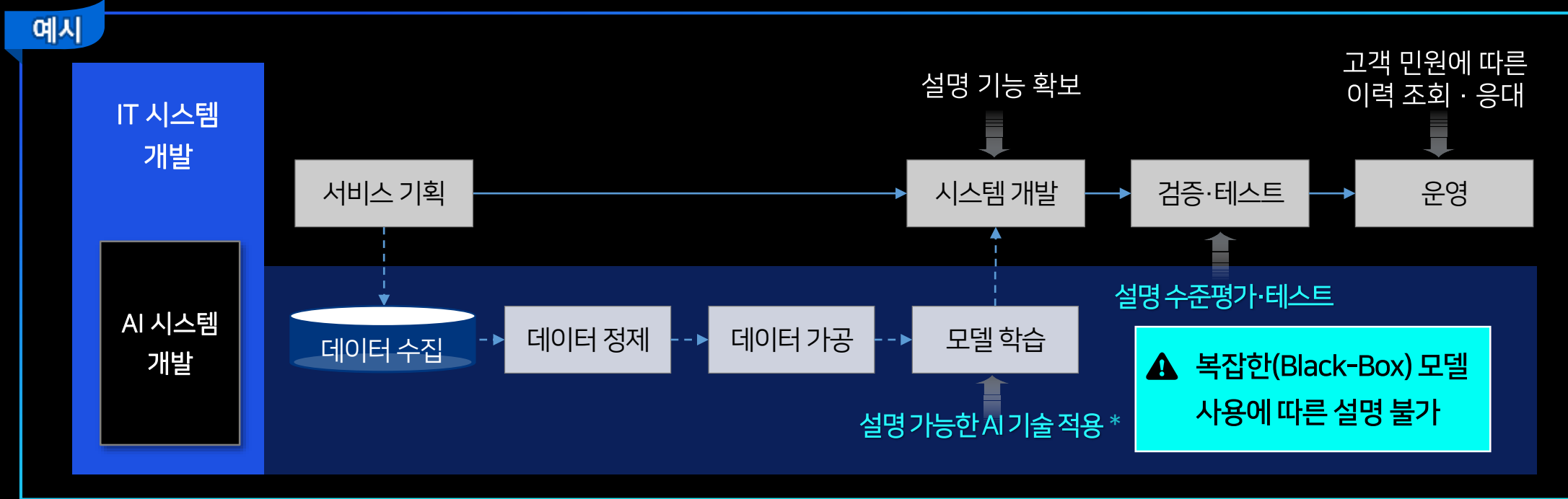
금융위원회 안내서	• 알고리즘 개발단계에서부터 편향을 완화, 공정성 수립기준을 만들어 평가하고 개선
신용정보법	• 합리적 이유 없이 성별 · 출신지역 · 국적 등으로 개인의 신용평점 차별 금지
금융소비자 보호법	• 금융소비자를 성별 · 학력 · 장애 · 사회적 신분 등을 이유로 부당하게 차별해서는 안됨

예시



고려사항 | ② 설명요구권 보장

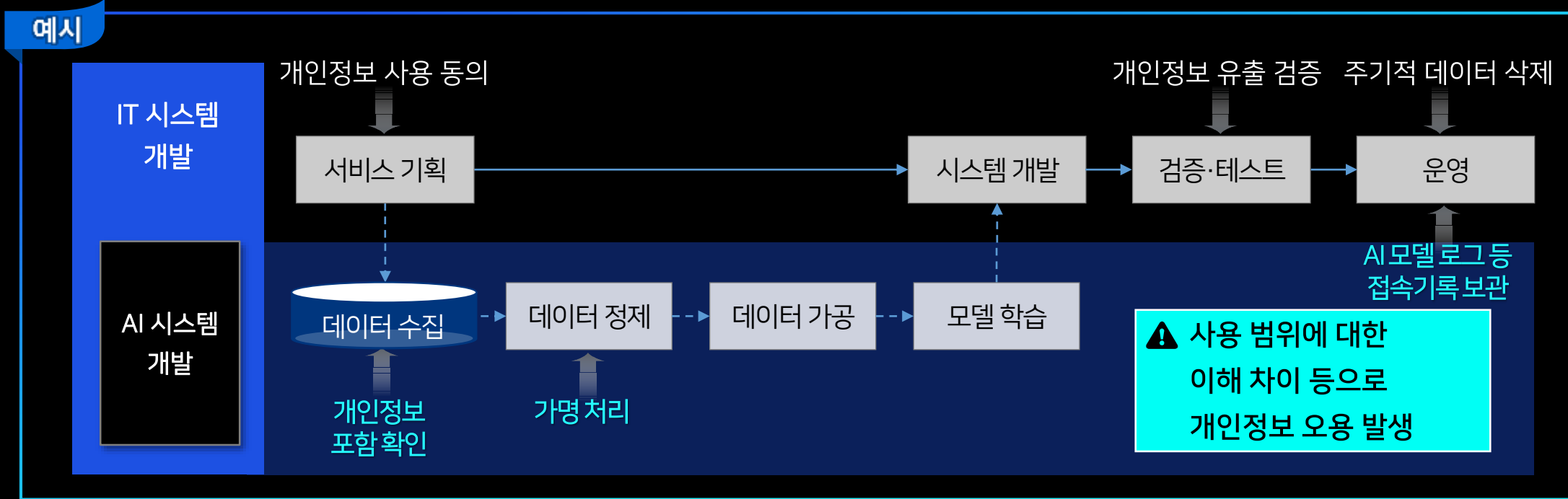
금융위원회 안내서	• 투명한 AI 기술을 적용하여 맥락에 맞게 설명하고 설명가능성을 합리적 수준으로 개선
신용정보법	• 신용정보주체는 자동화 평가 여부 · 결과 · 기준 · 이용정보에 대한 설명을 요구할 수 있음
개인정보 보호법	• 개인정보처리자가 자동화된 결정을 하는 경우, 설명요구에 필요한 조치를 취해야 함



* ref. www.samsungsds.com/kr/insights/explainable_ai.html

고려사항 | ③ 개인정보 보호

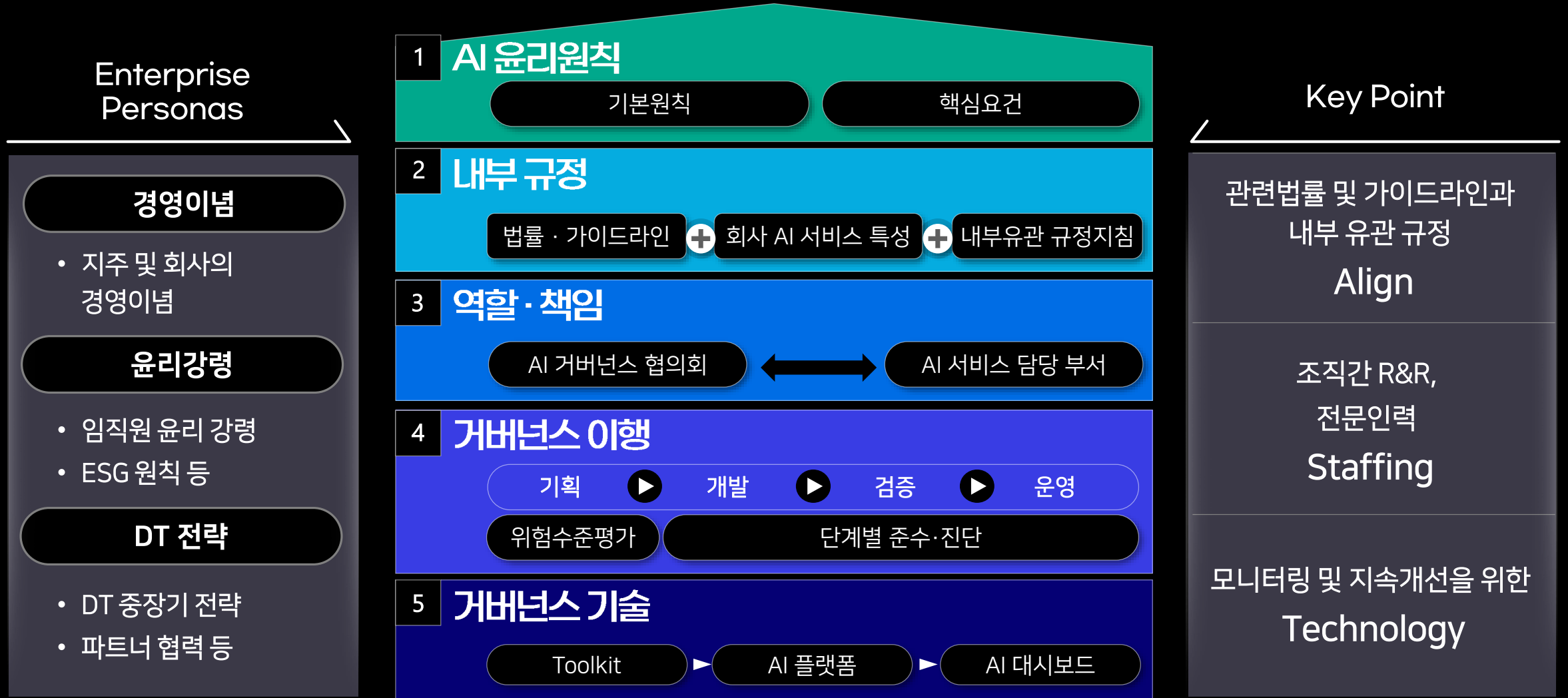
금융위원회 안내서	<ul style="list-style-type: none"> 해당 정보 활용의 필요성을 평가하고 사전 수집 동의 획득 및 비식별조치 등 안전 확보
개인정보 보호법	<ul style="list-style-type: none"> 내부 관리계획 수립, 접속기록 보관, 불필요 시 파기 등 기술적 · 관리적 · 물리적 조치



III.

거버넌스 수립방안

AI 거버넌스 추진 Framework

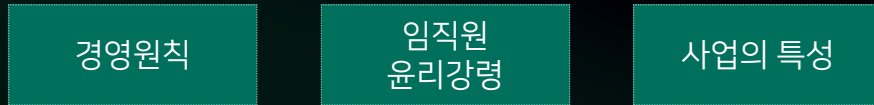


1 AI 윤리원칙

과학기술정보통신부 AI 핵심 요건 >>> >>>



“ 정부 AI 핵심 요건 및 기업의 특수성 고려 ”



기업의 내부 특성 >>> >>>

예시

AI 윤리원칙

> 인권 존중

삼성의 경영원칙을 바탕으로 인간 중심의 가치를 최우선, 다양한 사회적 가치와 국제인권법을 준수

> 다양성과 포용성

AI 기술의 적용에 평등과 다양성의 가치를 최우선, 정보사회에 소외되기 쉬운 취약계층의 고려

> 데이터와 프라이버시 보호

데이터 처리의 전 과정을 투명하게 관리, 사이버 공격으로부터 보안 프로세스를 강화

> 환경과 생태계 보전

생태계의 지속가능성에 부정적 영향 예방, 각국의 규범과 기준, 관행을 준수

> 커뮤니케이션

사용자에 대한 충분한 설명, AI의 위험과 부정적 결과에 대해 함께 대응 방안을 마련

2 거버넌스 내부 규정

“ 전체 정합성 유지 ”

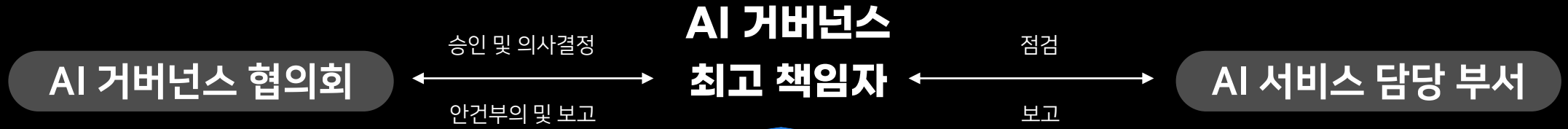
금융사 내부 유관 규정



관련 법률 및 가이드라인



3 거버넌스 역할 · 책임



사례

글로벌 금융사



- AI Ethics Leader
- AI 부문 총책임자



- RAI (Responsible AI) 프레임워크 팀 리더
- AI/ML 총책임자



- PURE* Data ethics framework 위원회 위원장
- Data/Transformation 총책임자



- Responsible AI 총책임자
- 외부인사
- * 퍼듀대 정치학과의 AI Policy & Ethics 교수



- Head of Data Ethics
- CPO (Chief Privacy Office)

* PURE : Purposeful, Unsurprising, Respectful and Explainable

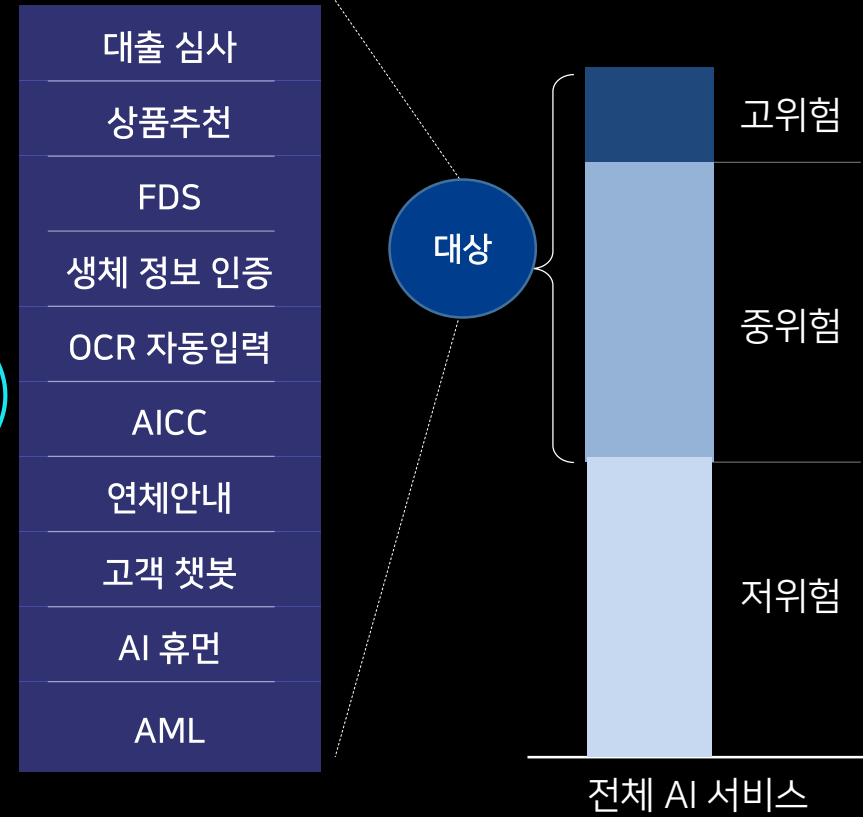
4 거버넌스 이행 | 위험수준 평가

예시

- 인권존중** "AI 서비스가 고객의 권리를 침해 하거나, 신체 / 재산적 피해를 발생시키지 않는가?"
- 포용성** "특정 집단이 AI 서비스 접근 시 차별적 요소가 존재하는가?"
- 프라이버시** "AI 활용 및 출력 정보에 개인정보 등 민감 데이터가 포함되어 있는가?"
- 환경** "AI 서비스가 개발, 운영 시 과도한 자원과 에너지를 사용하지 않는가?"
- 커뮤니케이션** "AI 서비스 모델이 이해가 어렵거나 블랙박스 형태로 관리 되지는 않는가?"

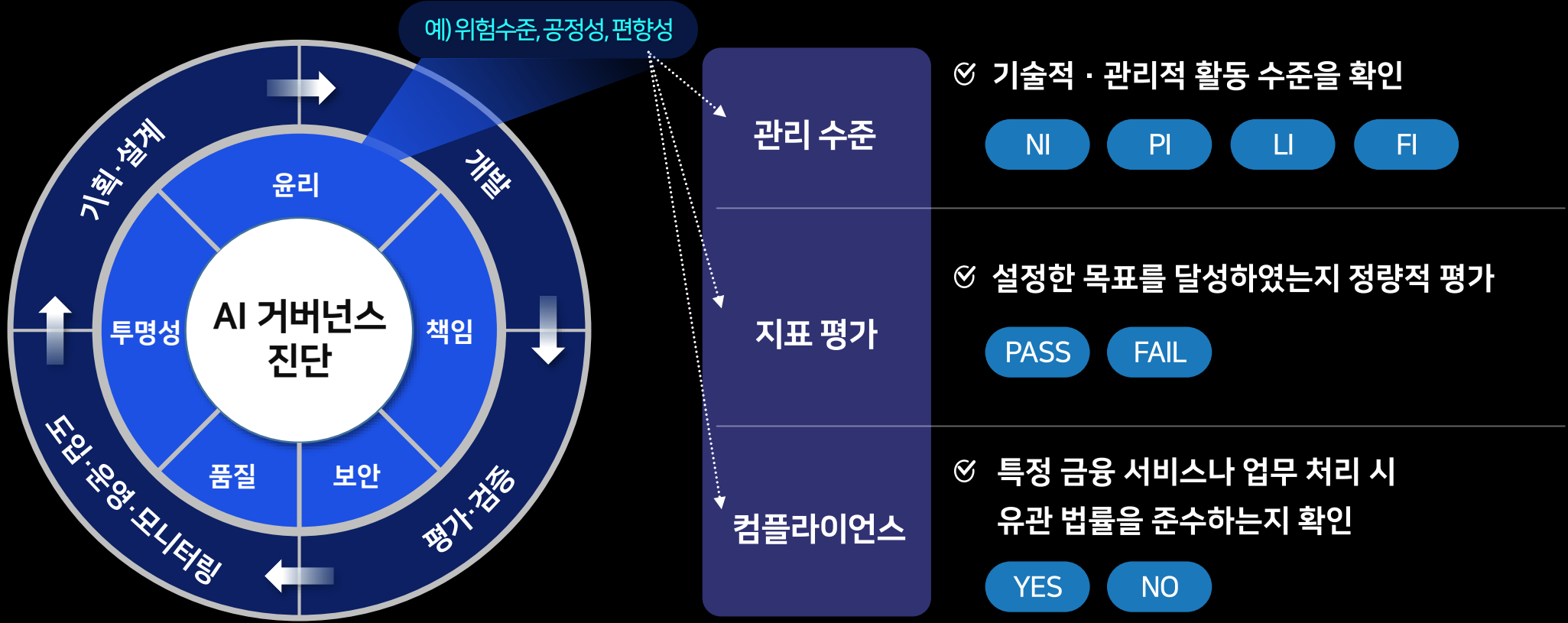
위험평가 방법

예시



4 거버넌스 이행 | 단계별 준수여부

삼성SDS 금융 AI 진단 Framework

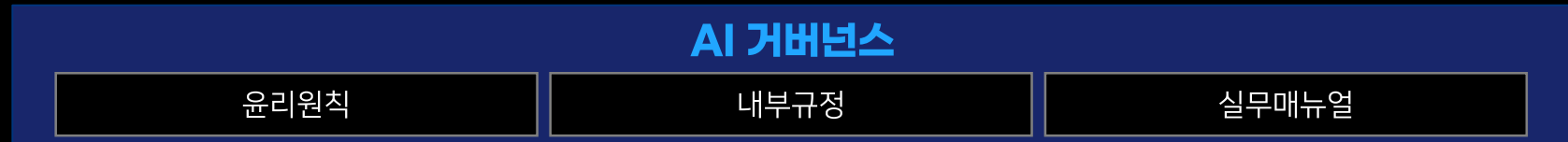


※ NI (Not Implemented), PI (Partially Implemented), LI (Largely Implemented), FI (Fully Implemented)

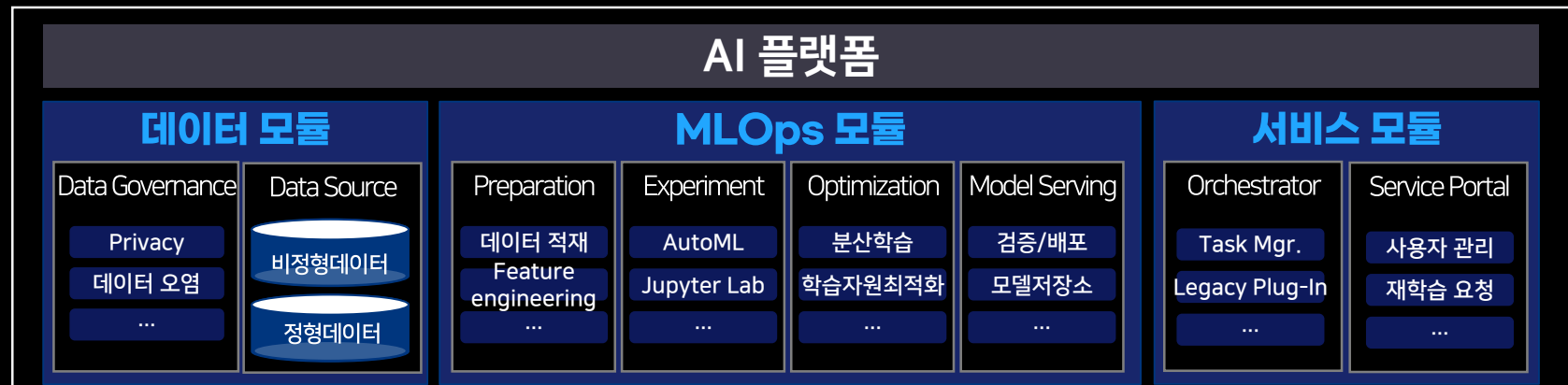
5 거버넌스 기술



AI 거버넌스 담당



AI 서비스 담당



AI 시스템 담당



위험은 줄이고 기회는 최대화 하기 위해...

> **우리회사 AI 서비스 중에 위험한 AI 서비스가 있는가?**

- AI의 정의와 위험에 대한 인식 재고
- AI 서비스 기획, 개발, 서비스 운영 담당자의 공감대

> **AI 거버넌스를 수립 한다면, 누가 주도해야 하는가?**

- 전문성, 경험 필요성, 이해상충 여부 판단
- 사내에 산재된 내부통제 부서의 역할 조정

> **AI 거버넌스를 시급하게 준비해야 할 이유는 있는가?**

- 불확실성과 리스크를 제거하고 활용에 집중
- 소비자 기대 수준 고려



Thank you

삼성SDS 조남용 프로
nyjo@samsung.com

SAMSUNG SDS