나를 대신해 일하는 AI(신원인증)

My AI, My ID:
The Future of Identity

이 AI 가 정말 나를 대신한다는 것을
어떻게 증명할까요 ?

국민대학교 · 윤승환 연구교수

내 삶 속으로 들어오는 AI: ICT 표준 R&D 신규과제 발표회



“부산 호텔 추천해줘”에서 “부산 휴가 갈래”로

검색 결과를 고르던 사람이, 이제는 AI에게 일을 맡기는 장면을 상상해 봅니다.

검색을 부탁할 때

AI가 후보를 추천합니다. 호텔 비교와 예약, 식당 대기 신청,
결제는 사람이 서비스마다 직접 처리합니다.

일을 맡길 때

AI가 취향에 맞는 호텔을 찾고, 예약과 맛집 웨이팅을
이어 처리합니다. 결제는 사용자가 정한 조건 안에서 진행합니다.

새롭게 필요한 확인

서비스는 요청을 보낸 AI가 누구인지, 누구를 대신하는지,
이번 행동까지 맡았는지를 확인할 수 있어야 합니다.

이 자료의 여행 이야기는 기술이 지향하는 활용 예시입니다.

휴가는 한 번의 부탁 , 실행은 여러 번의 확인

“나 다음 주에 부산 휴가 갈래 . 조용한 호텔로 찾아줘 .”

호텔을 예약할 때

AI 가 일정과 취향에 맞는 방을 찾습니다 . 예약 서비스는
이 AI 의 신원과 사용자가 맡긴 예약 범위를 확인합니다 .

맛집 대기 신청

AI 가 식당의 대기 가능 여부를 알아봅니다 . 식당에는
대기 신청에 필요한 정보와 권한만 전달합니다 .

내 카드로 결제할 때

결제 대상과 금액이 승인 범위에 드는지 확인합니다 .
추가 예약금이나 조건 변경은 사용자에게 다시 묻도록 설계합니다 .

편리함의 출발점은 “지금 요청하는 AI 가 누구인가”를 확인하는 일입니다 .

“AI 전용 주민등록증” 은 이해를 돕는 비유

AI 가 자신의 신원과 자격을 디지털 방식으로 증명한다는 뜻입니다 .

누구의 신원인가

사람의 신원과 일을 수행하는 AI 측 주체의 신원을 구분합니다 .
어떤 서비스 · 프로그램 · 실행 주체를 식별할지는 설계로 정합니다 .

무엇을 확인하나

믿을 수 있는 발급자가 만든 증명인지 , 유효한지 ,
그 증명이 실제 요청을 보낸 주체와 연결되는지 확인합니다 .

비유의 경계

실제 주민등록증이나 주민등록번호를 AI 에게 발급한다는 뜻은
아닙니다 . 하나의 카드나 단일 표준의 공식 명칭도 아닙니다 .

신원을 확인해도 , 그 AI 가 할 수 있는 일은 별도로 확인해야 합니다 .

이름표를 보여주는 것만으로는 부족합니다

호텔 직원이 이름표의 사진과 눈앞의 사람을 함께 살피듯 , 디지털 요청도 연결을 확인합니다 .

증명을 믿을 수 있나

발급자를 신뢰할 수 있는지 , 내용이 바뀌지 않았는지 ,
유효기간과 상태가 적절한지 검증합니다 .

지금 요청한 주체인가

해당 증명이나 토큰에 연결된 비밀키를 보유했음을 확인하는
방식 등을 사용해 , 복사한 증명만 내미는 위험을 줄입니다 .

이번 요청에 맞나

대상 서비스와 요청 맥락을 확인하고 재사용 공격을 막습니다 .
구체적인 검증 방법은 선택한 규격과 신뢰 체계에 따라 다릅니다 .

디지털 증명과 실제 요청 주체를 연결해야 사칭 위험을 줄일 수 있습니다 .

인증 , 위임 , 인가는 서로 다른 질문입니다

세 가지를 함께 확인해야 “나를 대신하는 AI 의 요청” 을 판단할 수 있습니다 .

인증
Authentication
“당신은 누구인가요 ?”
요청을 보낸 주체가 주장하는 그 주체인지 확인합니다 .

위임 Delegation
“누구를 대신해 무엇을 맡았나요 ?”
사용자가 AI 에게 맡긴 관계와 업무 범위를 확인합니다 .

인가 Authorization
“이 요청을 지금 허용할까요 ?”
서비스가 위임 범위와 자체 정책을 검토해 실행 여부를 결정합니다 .

AI 신원 확인이 출발점이고 , 위임과 인가가 실제 행동의 범위를 정합니다 .

20 만 원을 맡겼는데 23 만 원이 됐다면 ?

설명용 예시 : 사용자가 이번 호텔 예약에 1 회 , 20 만 원까지 결제를 허용했습니다 .

예약 요청 18 만 원

신원 · 위임 · 유효기간 등 필요한 검증을 통과하면 ,
결제 서비스가 자체 정책에 따라 승인 여부를 판단합니다 .

조건 변경 23 만 원

승인한 금액 범위를 넘습니다 . 자동 실행을 멈추고
변경된 조건에 대한 사용자의 추가 승인을 받도록 설계합니다 .

누군가 증명을 복사

증명을 제시했다는 이유만으로 결제하지 않습니다 .
실제 요청 주체와의 연결 및 재사용 여부를 함께 검증합니다 .

신원 인증은 사칭 방지의 기반입니다 . 결제 통제와 이상거래 탐지도 함께 필요합니다 .

호텔 · 식당 · 결제가 서로 다른 플랫폼이라면

사용자의 부탁은 하나지만 , 각 서비스가 신원과 자격을 해석하는 방식은 다를 수 있습니다 .

공통으로 읽는 방법

누가 발급한 어떤 증명인지 , 누구에 관한 정보인지 ,
무엇을 검증해야 하는지 이해할 공통 규칙이 필요합니다 .

필요한 만큼의 증명

호텔에는 예약에 필요한 증명을 , 식당에는 대기 신청에 필요한
증명을 제시합니다 . 모든 서비스에 같은 개인정보를 뿌리지 않습니다 .

각자의 허용 판단

상호연동은 무조건적인 신뢰나 자동 승인을 뜻하지 않습니다 .
서비스마다 신뢰할 발급자와 허용할 행동을 정해야 합니다 .

같은 AI 의 요청임을 확인하되 , 서비스에 필요한 자격과 권한을 구분합니다 .

맡긴 뒤에도 확인하고 멈출 수 있어야 합니다

신원 인증은 중요한 기반입니다 . 안심하고 쓰려면 서비스의 운영 장치도 함께 필요합니다 .

필요한 정보만 전달

여행 업무에 필요한 정보만 공개하도록 설계합니다 .
어떤 정보가 어느 서비스에 전달되는지 알 수 있어야 합니다 .

맡긴 일을 중단

위임의 만료와 철회를 다음 요청의 검증에 반영해야 합니다 .
이미 완료한 예약 · 결제의 취소는 해당 서비스의 별도 절차입니다 .

문제가 생기면 추적

누가 무엇을 승인했고 어떤 AI 가 무엇을 실행했는지 기록합니다 .
신원이 확인된 AI 도 실수할 수 있으므로 결과 확인이 필요합니다 .

“누구인지 안다” 는 사실만으로 모든 오류와 사기를 막을 수는 없습니다 .

이 과제는 플랫폼 사이의 증명 규칙을 다룹니다

AI Agent 기반 이기종 플랫폼 활용을 위한 신원 및 자격 증명 상호연동 프로토콜 표준 개발

쉽게 말하면

서로 다른 서비스가 AI 의 신원과 자격 증명을 주고받고 검증하는 방법을 맞추기 위한 표준 개발 과제입니다 .

일상에서 기대하는 변화

호텔과 식당 , 결제 서비스를 오가며 맡긴 일이 이어지도록 신뢰의 기반을 마련하는 것입니다 . 예약 · 결제 기능은 각 서비스의 연동과 정책도 함께 갖춰야 작동합니다 .

여행 사례와 참고 규격은 이해를 위한 설명이며 , 과제의 채택 규격 · 성과를 뜻하지 않습니다 .

이 과제의 두 최종 목표

신원·자격증명을 교환하는 프로토콜과, 그 신뢰를 평가·운영하는 KYA 프레임워크를 함께 개발합니다.

최종 목표 1

보안 규격 및 시험방법

- 이기종 플랫폼 간 AI Agent 신원·자격증명 교환
- 생성·저장·교환·검증·관리 절차와 메시지 구조
- 상호운용성·보안성 시험방법

최종 목표 2

KYA 프레임워크

- 발급자·소유자·실행 주체와 권한 위임 관계
- 신뢰 평가 모델과 정책 연계 메커니즘
- 신원 검증·자격/속성 확인·행위 기반 신뢰도

관련과제 | AI Agent 기반 이기종 플랫폼 활용을 위한 신원 및 자격 증명 상호연동 프로토콜 표준 개발

플랫폼이 달라도, 같은 신원과 자격을 확인해야 합니다

AI가 여러 회사를 오갈 때마다 신원이 끊기거나, 같은 증명을 서로 다르게 해석하면 자동 실행은 안전하지 않습니다.

신뢰 서비스 · 자격증명 발급 · 신뢰 앵커 · 상태 확인



사용자
부산 휴가를 맡김

내 AI Agent
요청·증명·위임 전달

호텔 서비스
신원·자격·정책 확인

결제 서비스
대상·금액·상태 확인

같은 의미
신원·자격 데이터 구조

같은 검증
서명·발급자·상태

같은 경계
대상·행위·기간·금액

같은 실패원칙
확인 불가 시 중단

서로 다른 플랫폼이 같은 증명의 뜻과 상태를 이해해야, 사용자의 위임이 끊기지 않습니다.

식별·인증·자격·위임·인가는 서로 다른 질문입니다

하나가 확인됐다고 나머지가 자동으로 따라오지 않습니다.

식별 Identification	어느 AI인가?
인증 Authentication	주장한 그 AI가 맞나?
자격 증명 Credential	어떤 속성·자격이 유효한가?
위임 Delegation	누구를 대신해 어디까지 하나?
인가 Authorization	이 요청을 지금 허용할까?

“진짜 AI”라고 확인해도 결제 권한이 생기는 것은 아닙니다.

서비스는 위임 범위와 자체 정책을 확인해 최종 인가를 결정합니다.

대리결제 사기를 줄이려면, 신원 확인만으로는 부족합니다

과제 기술은 “사기를 완전히 방지”한다고 보장하는 것이 아니라, 위조·오남용을 줄일 검증 기반을 제공합니다.

설명용 미래 시나리오

18만 원 → **23만 원**

허용 범위 / 한도 초과

AI가 “이 정도는 괜찮겠지”라고 넘어가는 것이 아니라,
결제 서비스가 요청을 거절하고 다시 확인하도록 설계합니다.

실행 전에 확인할 네 가지

요청 주체	실제 요청을 보낸 AI와 키가 증명에 연결되는가?
증명 상태	발급자·서명·유효기간·폐기 상태가 유효한가?
위임 범위	호텔·금액·기간·횟수·재위임 조건 안인가?
거래 상태	중복 요청·재사용·철회 후 요청이 아닌가?

확인할 수 없으면 실행하지 않는다.

필요한 정보만, 필요한 순간에 증명합니다

신원 확인을 위해 사용자의 모든 정보나 AI의 전체 활동 이력을 공개할 필요는 없습니다.

호텔 서비스의 질문

"이 AI가 20만 원 이하의 숙박 예약을 맡았습니까?"

→

최소 공개 증명

필요한 조건만 참/거짓으로 검증

→

검증 결과

조건 충족

주민번호·전체 여행 이력·불필요한 개인정보는 공개하지 않습니다.

ZKP	자격을 노출하지 않고 조건 충족을 증명
가명·익명화	멀티모달·행동 데이터의 재식별 위험을 감소
차등 프라이버시	통계·스트림 데이터 활용과 보호의 균형
이벤트 무결성	동적 AI의 신원 연속성과 위·변조를 검증

한 번 발급하고 끝나는 신원이 아닙니다

AI는 업데이트·복제·재시작될 수 있으므로, 현재 실행 주체와 상태가 계속 이어지는지 확인해야 합니다.



발급
신원·자격 발급

활성
정상 요청·검증

보류
위험·변경 시 중지

폐기
더 이상 신뢰하지 않음

감사
행위·위임 흐름 추적

이벤트 기반 서명·무결성

업데이트·이동·재시작 과정에서 신원 연속성을 검증

상태 천이

유효·보류·폐기 상태를 여러 플랫폼이 같은 의미로 확인

책임 추적성

누가 발급·소유·실행·위임했는지 감사 가능한 기록 확보

주의 | 권한을 폐기하는 것과 이미 완료된 예약·결제를 취소하는 것은 별도 업무입니다.

표준화 개발은 9대 핵심기술을 하나의 참조 구조로 연결합니다

연결에 성공하는 경우뿐 아니라, 거절·보류·폐기되어야 하는 경우까지 시험 가능한 구조를 지향합니다.

IDENTITY & PRIVACY

KYA 인증
발급자·소유자·실행 주체

가명·익명화
멀티모달 데이터 보호

차등 프라이버시
실시간 데이터 보호

TRUST & VERIFICATION

ZKP 자격 검증
최소 정보 공개

이벤트 기반 검증
신원 연속성·무결성

권한 위임·재위임
범위 확대 방지

AGENT & INFRASTRUCTURE

AI Agent DID/VC
식별자·자격 증명 관리

수명·상태 관리
유효·보류·폐기

분산원장 앵커
무결성·교환 근거

통합 참조 구현체 + 시나리오 데모 + 상호운용성·보안성 시험

계획서 기준 대상 표준화 기구: ISO/IEC JTC 1/SC 27, ITU-T SG17, IEEE 3079, TTA/KS · 개발·제안 계획이며 승인 성과를 의미하지 않음

이 과제가 바꾸려는 일상은 단순합니다

복잡한 기술은 뒤에서 작동하고, 사용자는 맡긴 일의 범위와 결과를 이해할 수 있어야 합니다.

지금		과제가 지향하는 미래	
검색	후보를 찾는다	**한마디로 맡긴다**	AI가 여러 서비스를 연결
확인	조건을 직접 비교한다	**확인 가능한 신원**	누가 보낸 어떤 AI인지 검증
예약	서비스마다 다시 로그인한다	**경계 안에서 실행**	범위를 넘으면 멈추고 재확인
결제	카드와 개인정보를 반복 입력한다	**결과와 책임 확인**	무엇을 했는지 추적·감사

표준은 모든 AI를 똑같이 만드는 규칙이 아니라,
서로 다른 AI와 서비스가 같은 신원·자격·경계를 이해하게 하는 약속입니다.

다음에 AI 에게 일을 맡길 때 기억할 질문

이 AI 는 누구인가요 ?
정말 나를 대신하나요 ?
내가 맡긴 일의 범위 안인가요 ?

“나를 대신하는 AI 가 누구인지 확인하고 ,
내가 맡긴 일이 무엇인지 함께 확인할 수 있어야 합니다 .”