AI 에이전트 시대의 디지털 신원 지갑 : 신원 증명에서 위임·승인·실행 검증으로

한국전자통신연구원
사이버보안연구본부
AI보안연구실

진승헌 책임연구원

초록

디지털 신원 지갑은 이용자의 신원과 자격을 증명하는 수단으로 발전해 왔다. 그러나 AI 에이전트가 이용자를 대신해 반복적으로 거래를 실행하는 환경에서는 신원뿐 아니라 위임의 범위, 승인의 내용, 실행의 정당성을 함께 검증해야 한다. NIST SP 800-63C-4의 가입자 통제형 지갑은 신뢰당사자에게 제시할 새로운 어서션의 서명에 가입자 활성화를 요구하고, EUDI Wallet ARF는 전자결제 승인 과정에서 거래 내용을 이용자에게 확인·승인받고, 승인된 거래 데이터를 기기 바인딩 서명에 반영한다. 반면 에이전트 기반 결제 프로토콜인 AP2는 사전 위임을 근거로 이용자가 실행 시점에 개입하지 않는 Human Not Present(HNP, 무인 실행)를 다룬다. 본고는 이들 체계 사이의 핵심 공백을 이용자의 승인, 제한된 실행 권한, 에이전트의 실행과 결과 증적을 검증 가능하게 이어 주는 책임사슬의 문제로 본다. 이를 위해 신원·권한·의도·에이전트·실행의 다섯 검증축과 신원·자격 확인–위임 승인–실행 인가–실행·재검증–결과 증적의 다섯 단계를 제시하고, 이용자 개입 수준을 직접 승인(Attended)·이용자 감독 실행(User-Supervised)·무인 실행(Unattended)으로 구분한다. 또한 RFC 9396의 RAR를 기반으로 AI 에이전트 위임의 의미·평가·연결 기준을 정의하는 정규화 인가 프로파일과, 이를 특정 위임에 적용한 정규화 인가 객체(NAO)를 제안한다. 마지막으로 국내 전자금융거래법의 접근매체 규율을 검토하고, 승인–인가–실행–증적을 연결하기 위한 향후 과제를 제시한다.



I 서론

디지털 신원 지갑의 기본 질문은 “이 이용자가 누구인가”였다. 이용자가 지갑을 활성화하고 필요한 신원·자격 정보를 선택해 제시하면 신뢰당사자(Relying Party, RP)가 이를 검증한다. 그러나 AI 에이전트가 이용자를 대신해 외부 서비스를 호출하고 거래를 수행하기 시작하면서 질문의 범위가 달라지고 있다.

가령 이용자가 AI 에이전트에게 “이번 달 사무용품을 총 50만 원까지 알아서 구매하라”고 지시했다고 하자. 이후 실제 구매는 이용자가 없는 시점에 여러 차례 이루어질 수 있다. 이때 신원이 확인됐다는 사실만으로는 충분하지 않다. ‘이번 달’의 범위는 무엇인지, 50만 원은 건별인지 누적인지, 어떤 판매자와 품목까지 허용했는지, 지금 요청하는 에이전트가 실제로 권한을 위임받은 에이전트인지, 승인 이후 권한이나 정책이 변경되지는 않았는지를 함께 확인해야 한다. 따라서 AI 에이전트 환경에서 새롭게 부각되는 문제는 거래별 이용자 승인의 부재라기보다, 사전 승인을 근거로 이후의 제한된 실행 권한을 어떻게 안전하게 구성하고 통제할 것인가이다. 본고는 이 문제를 이용자의 신원·자격 확인에서 위임 승인, 실행 인가, 실제 실행과 결과 증적까지를 검증 가능하게 이어 주는 책임사슬의 문제로 본다.

이를 위해 먼저 NIST 가입자 통제형 지갑의 이용자 활성화와 EUDI Wallet의 거래별 승인을 비교하고, AP2가 Human Not Present 실행을 어떻게 구성하는지 살펴본다. 이어 승인과 실행의 시간적 분리에서 발생하는 문제를 분석하고, RAR 기반 정규화 인가 프로파일과 NAO, 신원·권한·의도·에이전트·실행의 다섯 검증축과 위임·실행 책임사슬의 다섯 단계를 제시한다.

II 거래별 승인에서 무인 실행으로

1. NIST 가입자 통제형 지갑과 EUDI Wallet의 이용자 개입

NIST SP 800-63-4는 디지털 신원 보증을 신원확인 보증 수준(Identity Assurance Level, IAL), 인증 보증 수준(Authentication Assurance Level, AAL), 연합 보증 수준(Federation Assurance Level, FAL)으로 구분한다.

[1] 그러나 IAL·AAL·FAL은 각각 신원확인, 인증, 연합 거래의 보증 강도를 나타내는 기준이지, 특정 거래에서 이용자가 직접 내용을 확인·승인했는지, 에이전트가 이용자 감독하에 실행했는지, 또는 무인 실행되었는지를 표시하는 기준은 아니다. 이와 관련하여 SP 800-63C-4의 가입자 통제형 지갑(subscriber-controlled wallet) 모델은 <그림 1>과 같이 가입자 통제형 지갑이 RP에게 제시할 어서션[1]에 서명할 때에는 가입자의 활성화 요소(activation factor)가 요구된다. 지갑이 동일한 RP에 대한 이용자의 이전 정보공개 또는 인가 결정을 저장하고 있더라도, 새로운 어서션에 서명할 때 필요한 활성화 절차는 생략할 수 없다.[2]

그림 1 가입자 통제형 지갑(Subscriber-Controlled Wallet) 기반의 연합 디지털 신원 모델 [1]

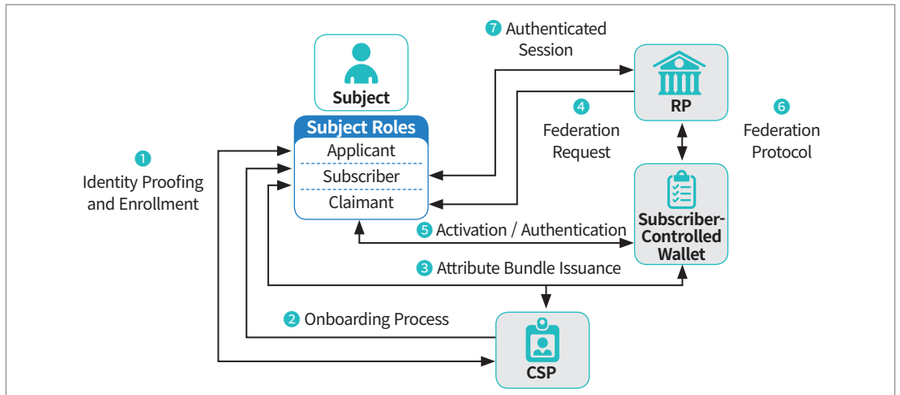
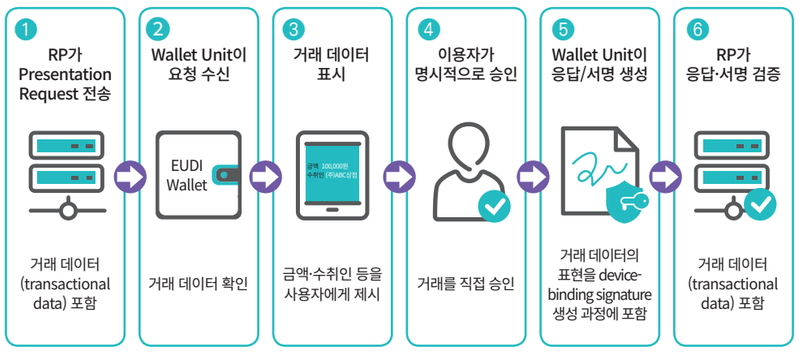


그림 2 전자결제 SCA를 위한 EUDI Wallet의 거래별 승인·서명 흐름



출처: EUDI Wallet ARF 3.0.0을 바탕으로 저자 재구성

1) 어서션(assertion)은 가입자의 인증 사건에 관한 정보를 신원제공자(IdP)가 RP에게 전달하는 전자적 진술로, 필요에 따라 가입자의 신원 속성도 포함할 수 있다.

EUDI Wallet도 신원·속성 제시에만 머물지 않는다. Regulation (EU) 2024/1183은 지갑을 이용자 통제하에 두고, 지갑을 통한 행위에 대해 안전하고 명시적·능동적인 이용자 확인을 요구하는 원칙을 제시한다.[3] ARF는 Regulation과 이를 구체화하는 집행위원회 시행규정(Implementing Regulations)을 바탕으로 EUDI Wallet의 기술적 구현과 상호운용성을 구체화하는 기술 참조 프레임워크이다. <그림 2>와 같이 ARF 3.0.0은 전자결제의 강한 고객인증(Strong Customer Authentication, SCA)을 지갑으로 지원하기 위한 강한 이용자 인증(Strong User Authentication, SUA) 시나리오를 제시한다. 이 시나리오에서는 제시 요청(presentation request)에 거래 데이터(transaction data)가 포함되며, 결제의 경우 적어도 금액과 수취인이 이에 해당한다. Wallet Unit은 거래 데이터를 이용자에게 표시해 승인을 받은 뒤, 승인된 거래 데이터가 기기 바인딩 서명(device-binding signature)에 반영되도록 한다. 이는 EU 지급서비스지침2(Payment Services Directive 2, PSD2)에서 요구하는 동적 연계(dynamic linking), 즉 인증 결과가 특정 거래의 금액과 수취인에 결합되어야 한다는 요구에 대응하도록, 이용자가 확인·승인한 거래 데이터가 지갑의 기기 바인딩 서명에 포함되게 한 구조로 볼 수 있다.[4]

또한 ARF 3.0.0은 부모·후견인나 위임장 등에 따른 자연인 간 대리관계를 대리관계 속성증명(representation attestation)으로 다루지만,[4] 이는 소프트웨어 에이전트가 사전 위임을 근거로 반복적으로 권한을 행사하는 구조와는 범위가 다르다.

NIST의 가입자 통제형 지갑과 EUDI Wallet의 전자결제 흐름은 모두 이용자의 직접 개입을 요구하지만, 그 개입이 적용되는 지점은 다르다. 요컨대 NIST는 RP에 제시할 새로운 어서션의 서명에 이용자 활성화를 요구하고, EUDI Wallet은 전자결제에서 현재 거래 내용을 이용자가 확인·승인하도록 한다. AI 에이전트 환경의 새로운 질문은 이러한 현재 시점의 개입을 넘어, 사전 승인을 미래의 제한된 반복 실행 권한으로 어떻게 안전하게 연결할 것인가이다.

2. AP2: 사전 위임에서 구체적 거래 인가로

에이전트 기반 결제 프로토콜인 AP2는 이용자가 실행 시점에 참여하는 Human Present와, 사전 위임을 근거로 이용자가 실행 시점에 직접 개입하지 않는 Human Not Present(HNP, 무인 실행)를 구분한다. HNP에서는 이용자가 미래 거래의 범위와 제약조건을 담은 사전 만다트(Open Mandate)를 승인하고, 이후 실제 거래 단계에서 상품·판매자·금액 등이 구체화된 확정 만다트(Closed Mandate)가 생성된다. 검증자는 사전 만다트와 확정 만다트의 연결관계, 사전 제약조건 충족 여부, 에이전트 키 보유 여부, 처리 증적(Mandate Receipt)을 확인한다. AP2는 이처럼 사전 위임–구체적 거래 인가–처리 증적을 결제 영역에서 연결하지만, 이를 신원 보증·실행환경 신뢰·법적 책임까지 포괄하는 범용 체계로 확장하지는 않는다.[5]

3. 기존 인가 기술은 어디까지 활용할 수 있는가

AI 에이전트 위임을 위해 완전히 새로운 인가 기술을 만들 필요는 없다. RFC 9396(Rich Authorization Requests, RAR)는 authorization_details 파라미터를 이용해 단순한 OAuth scope보다 세밀한 권한을 구조화하여 표현할 수 있고, 이를 이용자 동의와 액세스 토큰 발급에 연결한다.[6] 따라서 대상, 행위, 금액, 기간, 수신자와 같은 위임 조건의 기본 표현은 RAR를 활용하는 것이 합리적이다. 또한 RFC 9449(Demonstrating

Proof of Possession, DPoP)는 액세스 토큰을 특정 공개키에 바인딩하여 토큰 탈취 후 재사용 위험을 줄일 수 있다.[7] RFC 9470(OAuth 2.0 Step Up Authentication Challenge Protocol)은 자원 서버가 기존 인증의 강도나 최근성이 부족하다고 판단할 때 더 강하거나 더 최근의 인증을 요구할 수 있도록 한다.[8] 정리하면 개별 기술 요소의 상당수는 이미 존재한다. 부족한 것은 이용자의 승인, 그 승인을 근거로 부여된 제한된 실행 권한, 권한을 행사하는 에이전트 키, 구체적 실행과 결과 증적이 동일한 승인 근거를 추적 가능하게 참조하도록 하는 공통 연결 기준이다.

표 1 AI 에이전트 위임 관련 기존 체계 비교

체계	주로 해결하는 문제	AI 에이전트 환경에서 남는 문제
NIST SP 800-63C-4	이용자 활성화와 지갑 어서션	현재 이용자 활성화와 반복 실행 권한의 연결 기준
EUDI Wallet	신원·자격, 거래별 이용자 승인	미래의 반복적 무인 실행 권한
AP2	Open→Closed Mandate, HNP, Receipt 기반 상태 통제	결제 외 영역 및 신원·자격·실행환경·법적 책임과의 연결
OAuth RAR·DPoP	세밀한 인가 표현, 토큰–키 바인딩	승인·상태·실행·증적을 잇는 공통 프로파일

OAuth 계열 외에도 RFC 9635(Grant Negotiation and Authorization Protocol, GNAP)은 소프트웨어 클라이언트의 위임 인가와 권한 협상을 지원하지만,[9] 본고가 주목하는 승인 의미–상태 재검증–실행–결과 증적의 공통 책임사슬까지 정의하는 것은 아니다. 한편 NIST NCCoE(National Cybersecurity Center of Excellence)가 2026년 AI 에이전트 신원·인가 실증 방향에 대한 개념보고서를 공개한 것은 이 문제가 표준 기반 실증 과제로 구체화되고 있음을 보여준다.[10]

승인과 실행 사이의 구조적 공백과 대응 개념

1. 승인과 실행의 시간적 분리

승인과 후속 실행의 시간적 분리 자체는 기존 자동화 거래에도 존재한다. 그러나 AI 에이전트는 사전 승인을 근거로 실행 대상과 시점을 동적으로 선택하며 여러 차례 실행할 수 있으므로, 그 사이 이용자의 자격·계정 상태, 에이전트 키, 잔여 한도와 정책 등이 달라질 수 있다.

따라서 무인 실행에서는 승인 당시의 정당성과 실행 당시의 정당성을 구분해야 한다. 실행 시점에는 잔여 한도·사용 횟수·철회 여부 등 인가 상태(Authorization State)와 토큰·자격·에이전트 상태·위험 신호 등 실행 시점 문맥(Runtime Context)을 다시 평가해야 한다.

2. RAR 기반 정규화 인가 프로파일과 NAO

이용자가 승인한 권한을 AI 에이전트가 반복적으로 실행하려면 자연어 위임을 기계가 일관되게 해석·평가할 수 있는 조건으로 구체화해야 한다. 예를 들어 “이번 달 사무용품을 50만 원까지 구매하라”는 지시는

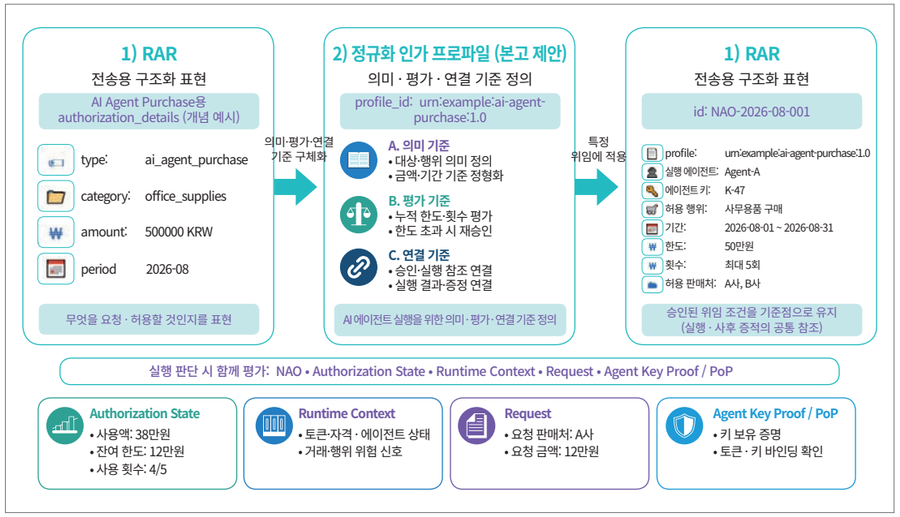
기간, 금액의 건별·누적 여부, 허용 품목과 판매자 범위 등을 명확히 정의할 필요가 있다.

RFC 9396의 세분화된 인가 요청(Rich Authorization Requests, RAR)은 authorization_details 파라미터를 이용해 이러한 세밀한 인가 조건을 구조화하여 표현할 수 있는 기반을 제공한다.[6] 다만 RFC 9396은 authorization_details 파라미터를 통해 세밀한 인가 조건을 표현할 수 있는 틀을 제공할 뿐, 금액·기간·대상·횟수와 같은 조건을 서비스 간에 동일한 의미로 해석하거나, 특정 요청이 승인된 위임 범위 안에 있는지를 판단하는 공통 규칙까지 정의하지는 않는다.[6] 따라서 AI 에이전트 위임에서는 동일한 위임 조건을 일관되게 해석·평가하기 위한 별도의 정규화 인가 프로파일이 필요하다.

이에 본고는 AI 에이전트의 반복 실행에 필요한 의미·평가·연결 기준을 정규화 인가 프로파일(Normalized Authorization Profile)로 제안한다. 의미 기준은 금액·기간·대상·행위가 무엇을 뜻하는지를, 평가 기준은 누적 한도·횟수·재승인 조건을, 연결 기준은 이용자 승인과 실행 권한·에이전트 키·결과 증적의 연계 방식을 정한다. 여기서 ‘정규화’는 특정 직렬화 형식의 정규화를 의미하는 것이 아니라, 위임 조건의 의미와 평가 방법, 승인·실행·증적의 연결 방식을 일관되게 구체화한다는 뜻이다. 또한 본고에서는 이 정규화 인가 프로파일과 적용 정책을 특정 위임에 적용하여 구성한 논리적 인가 기준 객체를 정규화 인가 객체(Normalized Authorization Object, NAO)라고 부른다. NAO는 RFC 9396에 정의된 표준 객체가 아니라, 본고가 AI 에이전트 위임의 승인 범위와 실행 기준을 설명하기 위해 제안하는 개념이다.

NAO에는 대상·행위·금액·기간·횟수, 에이전트와 키 참조, 재승인 조건 등이 포함될 수 있다. 이때 RAR로 표현된 요청 조건에는 정규화 인가 프로파일의 의미·평가 기준과 적용 정책에 따른 추가 제한조건을 반영할 수 있으며, 이렇게 구성된 조건은 이용자에게 제시·승인된 후 해당 위임의 공통 인가 기준으로 확정된다. 승인 증거에는 승인 주체, NAO 식별자, 승인 당시 NAO 내용의 무결성 참조가 포함되어야 한다. 승인 증거의 신뢰 경로와 위임 상태 관리는 IV.4에서 별도로 다룬다.

그림 3 RAR, 정규화 인가 프로파일, NAO의 관계



NAO는 이용자가 승인한 범위와 조건을 기준값으로 담는다. 반면 실행에 따라 달라지는 현재 사용액·잔여 횟수·진행 중인 실행 등은 별도의 인가 상태로 관리한다. 따라서 실제 실행 시에는 NAO, 인가 상태, 실행 시점 문맥, 현재 요청(Request), 에이전트 키 보유 증명(proof of possession, PoP)을 함께 평가해야 한다. 실행 시점의 정책이나 위험 평가는 NAO에서 승인된 범위를 추가로 제한하거나 재승인을 요구할 수 있지만, 승인된 권한의 범위를 확대해서는 안 된다.

결국 RAR, 정규화 인가 프로파일, NAO는 각각 '구조화된 인가 표현–의미·평가·연결 기준–특정 위임에 적용된 인가 기준 객체'의 관계이다. 이 구조의 목적은 새로운 인가 프로토콜을 추가하는 것이 아니라, 이용자가 승인한 권한의 의미를 사전 승인에서 반복 실행과 결과 증적까지 일관되게 유지하고 검증 가능하게 연결하는 데 있다.

3. 이용자 개입 수준

기존 Human-in/on/out-of-the-Loop 개념을 참조하되,[11] 본고는 위임된 권한의 실행 시점에 이용자가 직접 승인·감독하는 정도에 초점을 맞춰 Attended, User-Supervised, Unattended의 세 수준으로 구분한다. 이는 새로운 보증수준을 제안하는 것이 아니라, 위임된 권한의 실행 시점에서 이용자의 개입 정도를 비교하기 위한 분류이다.

표 2 이용자 개입 수준

수준	의미	주요 확인·증거
Attended (직접 승인)	이용자가 현재 거래 내용을 직접 확인하고 승인	현재 거래 내용에 결합된 이용자 승인 증거
User-Supervised (이용자 감독 실행)	에이전트가 자동 실행하되 이용자가 실행 상태를 감독하고 필요 시 개입·중단 가능	감독 상태 또는 채널, 개입·중단 기능 및 관련 기록
Unattended (무인 실행)	사전 위임 범위에서 실행 시점에 이용자의 직접 개입 없이 에이전트가 실행	사전 위임·승인 증거, 제한된 실행 권한, 실행 시점 재검증 및 실행 결과 증적

EUDI Wallet의 전자결제 승인 흐름은 이용자가 현재 거래 데이터를 확인하고 승인한다는 점에서 Attended(직접 승인) 상태에 해당한다. 본고의 핵심은 이러한 직접 승인 결과를 근거로 이후 User-Supervised(이용자 감독 실행) 또는 Unattended(무인 실행)에 필요한 제한된 실행 권한을 어떻게 구성·통제할 것인가이다.

Ⅳ 위임 책임성을 위한 참조 모델

1. 다섯 검증축: 무엇을 확인해야 하는가

AI 에이전트가 수행한 거래의 정당성을 설명하려면 최소한 다음 다섯 가지 질문에 답할 수 있어야 한다. 각 질문은 서로 다른 검증 대상을 가지므로, 어느 하나가 충족되었다고 해서 다른 질문이 자동으로 해결되는 것은 아니다. 예를 들어 신원 확인이 강하더라도 위임 범위가 과도하면 실행의 정당성은 보장되지 않는다. 또한 에이전트가 특정 키를 보유하고 있다는 사실만으로 해당 소프트웨어·모델·실행환경까지 신뢰할 수 있다고 볼 수도 없다.

표 3 다섯 검증축과 핵심 질문

검증축	핵심 질문
신원	권한을 위임한 자연인 또는 법인은 누구인가
권한	무엇을 어디까지, 언제까지 위임했는가
의도	이용자에게 어떤 위임 내용이 제시·승인되었으며, 이후 실행이 그 승인 내용과 일치하는가
에이전트	실행 주체가 승인된 에이전트인가
실행	실제로 무엇이 어떤 조건에서 수행됐는가

여기서 '의도'는 이용자의 내면적 의도 자체를 추론한다는 의미가 아니라, 이용자에게 제시·승인된 위임 내용과 이후 실행의 일치성을 검증한다는 의미이다. 에이전트 검증에서는 키 보유와 실행환경 신뢰를 구별해야 한다. DPoP와 같은 기술은 특정 개인키의 보유를 증명하고 토큰을 해당 키에 바인딩할 수 있지만,[7] 에이전트 운영자, 모델·도구·정책의 버전, 실행환경의 무결성과 생애주기를 통합적으로 보증하지는 않는다.

2. 위임·실행 책임사슬의 다섯 단계

다섯 검증축이 "무엇을 확인해야 하는가"에 관한 기준이라면, 위임·실행 책임사슬의 다섯 단계는 그 확인이 거래 과정에서 어떤 순서로 이루어지고 어떤 증거로 남는지를 보여준다. 두 체계가 일대일로 대응하는 것은 아니다. 예를 들어 에이전트 검증은 실행 인가 단계에서 한 번 확인되는 데 그치지 않고, 실제 실행·재검증 단계에서도 계속 확인되어야 한다.

본고는 AI 에이전트의 위임과 실행 과정을 ① 신원·자격 확인, ② 위임 승인, ③ 실행 인가, ④ 실행·재검증, ⑤ 결과 증적의 다섯 단계로 구분한다. 중요한 것은 각 단계를 따로 확인하는 데 그치지 않는 것이다. 누가 어떤 자격으로 무엇을 승인했는지, 그 승인에 따라 어떤 실행 권한이 부여되었는지, 어느 에이전트가 그 권한을 행사했는지, 실제 결과가 무엇이었는지를 사후에도 연결해 검증할 수 있어야 한다. 특히 ④ 실행·재검증 단계에서는 승인 당시의 조건만 볼 것이 아니라, 실행 시점의 상태를 다시 평가해야 한다. 잔여 한도·사용 횟수·철회 여부 등 인가 상태와 토큰·자격·에이전트 상태·위험 신호 등 실행 시점 문맥을 함께 확인해야 하는 이유가 여기에 있다.

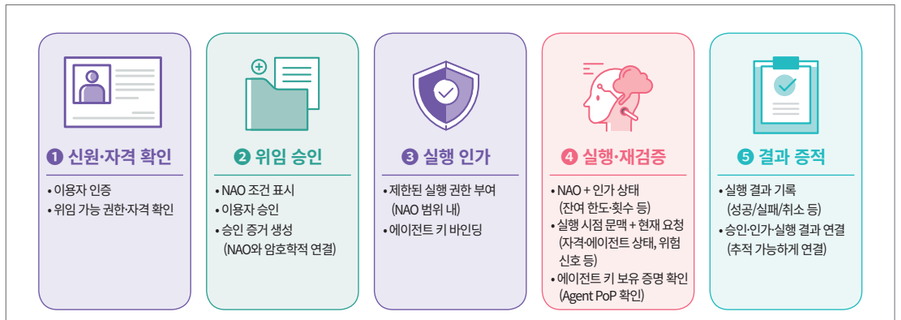
3. 사전 위임에서 무인 실행까지: 사례 분석

이 절에서는 "이번 달 지정된 판매처에서 사무용품을 총 50만 원까지 구매하라"는 위임 사례에 앞의 참조 모델을 적용해 본다. 이 지시는 자연어로는 간단하지만, 실제 실행을 위해서는 허용 품목, 판매처, 총 한도, 유효기간, 거래 횟수, 재승인 조건 등을 명확히 정해야 한다. 이러한 조건은 정규화 인가 프로파일과 적용 정책에 따라 NAO로 구성되고, 이용자의 확인·승인을 거쳐 해당 위임의 공통 인가 기준으로 확정된다.

이용자의 승인이 완료되면 인가 서버는 승인된 NAO를 근거로 에이전트에게 범위·기간·금액이 제한된 실행 권한을 부여한다. 예를 들어 에이전트가 지정된 판매처에서 12만 원의 사무용품 구매를 요청하면, 자원 서버와 정책 집행 구성요소는 해당 요청이 NAO의 승인 범위 안에 있는지, 인가 상태와 실행 시점 문맥이 실행을 허용하는지, 지정된 에이전트 키의 보유가 증명되는지를 함께 평가한다.

실행 결과는 실행 권한, NAO, 승인 증거와 연결해 기록하고 인가 상태에 반영해야 한다. 누적 한도나 거래 횟수 제한이 있는 경우에는 동시에 발생한 여러 실행이 동일한 잔여 권한을 중복 사용하지 않도록 진행 중인 실행도 상태에 반영해야 한다. 이 구조를 통해 Unattended 환경에서도 누가 무엇을 승인했고, 어떤 실행 권한이 부여되었으며, 어느 에이전트가 이를 행사해 어떤 결과를 만들었는지를 사후에 재구성할 수 있다.

그림 4 사전 위임에서 AI 에이전트의 무인 실행까지의 책임사슬



4. 승인 증거의 신뢰 경로와 위임 상태 관리

사전 위임의 책임성을 확보하려면 승인 증거가 단순히 존재하는 것만으로는 충분하지 않다. 검증자는 신원·자격이 확인된 이용자가 특정 NAO의 조건을 실제로 확인하고 승인했는지를 확인할 수 있어야 한다. 이를 위해 승인 증거에는 승인 주체에 대한 참조와 NAO 식별자가 포함되어야 하며, 승인 당시 NAO의 내용이 이후 변경되지 않았음을 확인할 수 있도록 NAO 내용의 해시 등 무결성 참조가 함께 남아야 한다.

본고에서 말하는 승인 증거(approval evidence)는 특정 표준의 고정된 객체명이 아니라, 이용자가 특정 NAO를 확인하고 승인했다는 사실을 나중에 검증할 수 있도록 남긴 기록을 뜻한다. 따라서 승인 증거는 반드시 별도의 표준 객체나 토큰 형태로 정의될 필요는 없다. 지갑이나 신뢰된 승인 환경(Trusted Surface)에서 생성된 서명 기록으로도 구현할 수 있다. 중요한 것은 그 형식이 아니라, 해당 기록이 신원·자격이 확인된 이용자의 실제 승인에 의해 생성되었음을 검증자가 신뢰 가능한 경로로 확인할 수 있어야 한다는 점이다. 이러한 신뢰 경로는 여러 방식으로 구성될 수 있다. 예를 들어 이용자 키에 대한 직접 검증, 자격증명 발급자가 보증하는 지갑 또는 신뢰된 승인 환경, 검증자가 신뢰하는 에이전트 제공자(Agent Provider)를 통한 확인 등이 가능하다.[5] 다만 어느 방식을 택하더라도 핵심은 동일하다. 어느 방식을 택하더라도 승인 증거는 특정 이용자, 특정 NAO, 특정 승인 시점을 함께 가리킬 수 있어야 한다.

이때 승인 증거, NAO, 인가 상태, 실행 시점 문맥의 역할을 구분해야 한다. 승인 증거는 "누가 어떤 NAO를 승인했는가"를 보여 준다. NAO는 "무엇을 어떤 에이전트에게 어디까지 허용했는가"를 나타내는 기준이다. 반면 인가 상태(Authorization State)는 "그 권한이 현재 얼마나 사용되었고, 남은 한도나 횟수는 얼마이며, 철회되었는가"를 관리한다. 토큰·자격·에이전트 상태와 위험 신호처럼 실행 시점에 달라질 수 있는 요소는 실행 시점 문맥(Runtime Context)으로 평가한다.

따라서 사용액, 잔여 횟수, 진행 중인 실행, 위임 철회와 같은 동적인 변화는 NAO를 직접 수정하기보다 인가 상태에 반영하는 것이 적절하다. NAO에 정의된 유효기간은 승인된 조건으로 유지하되, 실제 실행 시 현재 시각과 비교하여 유효 여부를 판단한다. 즉 NAO는 승인된 범위와 조건의 기준점이고, 인가 상태는 그 기준이 현재 얼마나 소비되었는지를 나타내며, 실행 시점 문맥은 지금 이 실행을 허용해도 되는지를 판단하는 데 사용된다.

승인된 위임의 범위를 확대하거나 중요한 조건을 변경해야 하는 경우에는 기존 NAO를 수정하기보다 새로운 식별자를 가진 NAO를 구성하고 필요한 재승인을 받아야 한다. 기존 NAO는 이전 승인과 실행의 사후 검증을 위해 참조 가능하게 유지한다. 반대로 위임 철회, 한도·횟수의 소진, 진행 중 실행과 같은 변화는 인가 상태에 반영하면 된다. 이러한 구분을 통해 최초의 이용자 승인, 제한된 실행 권한, 반복 실행, 결과 증적을 하나의 검증 가능한 책임사슬로 유지할 수 있다. 최초 승인은 이후의 모든 실행을 자동으로 정당화하는 것이 아니라, 제한된 실행 권한을 만들기 위한 기준점이다. 각 실행 시점에는 NAO의 승인 범위, 인가 상태, 실행 시점 문맥을 다시 평가해야 하며, 그 결과가 결과 증적과 연결되어야 한다.

Ⅴ 국내 법제와 정책적 고려사항

AI 에이전트 기반 금융거래 위임은 단순한 기술 설계 문제가 아니라 국내 전자금융거래 법제와 함께 검토되어야 한다. 특히 핵심 쟁점은 이용자의 기존 인증수단이나 접근매체를 에이전트가 직접 사용하는 구조를 허용할 것인지, 아니면 이용자 인증수단과 분리된 제한적 실행 권한을 별도로 부여할 것인지에 있다. 본고의 관점에서는 후자의 구조, 즉 이용자 접근매체와 에이전트 실행 권한의 분리가 더 안전하고 법제 친화적인 방향이다.

1. 이용자 접근매체와 에이전트 실행 권한의 분리

전자금융거래법은 금융회사 등이 전자금융거래에서 이용자의 신원, 권한 및 거래지시의 내용 등을 확인하도록 하고, 접근매체의 양도·대여·보관·전달 등을 엄격히 제한한다.[12] 대법원은 접근매체의 확정적 양도와 일시적 사용 위임을 구별하지만,[13] 시행령은 접근매체 사용 위임을 이용자의 중대한 과실 유형으로 규정한다.[14] 이러한 점에서 AI 에이전트에게 이용자의 기존 접근매체 자체를 맡기는 방식은 신중할 필요가 있다.

이용자의 접근매체는 신원·자격 확인과 위임 승인에 사용하고, 에이전트에게는 목적·대상·금액·기간이 제한된 별도의 실행 권한을 부여하는 구조가 바람직하다. 다만 에이전트 전용 키나 토큰이 전자금융거래법상 접근매체에 해당하는지는 구체적인 발급·등록·사용 방식에 따라 별도로 검토해야 한다.

2. 향후 과제

국내 디지털 신원 지갑이 AI 에이전트 환경으로 확장되기 위해서는 다음 세 가지 과제가 중요하다.

첫째, 승인–인가–실행–증적을 잇는 책임사슬의 연결 기준을 표준화해야 한다. 본고가 검토한 NIST SP 800-63C-4, EUDI Wallet, AP2, OAuth RAR·DPoP 등은 각각 신원 확인, 이용자 승인, 세밀한 인가 표현, 키 바인딩과 같은 요소를 제공한다. 특히 AP2는 결제 영역에서 사전 위임, 구체적 거래 인가, 처리 증적을 검증 가능하게 연결한다. 그러나 이를 결제 영역을 넘어 신원·자격, 범용 실행 인가, 에이전트 실행환경, 사후 증적까지 포괄하는 공통 책임사슬로 일반화한 기준은 아직 충분히 정립되어 있지 않다. 따라서 새로운 인가 프로토콜을 추가하기보다 RAR와 기존 지갑·OAuth 기술을 활용하면서, 위임 승인부터 실행 인가, 실행 시점 재검증, 결과 증적까지 필요한 참조·바인딩·추적 기준을 프로파일화하는 것이 현실적이다.

둘째, 보증 수준을 하나의 등급으로 단순화하지 말고, 여러 보완적 축으로 나누어 평가해야 한다. IAL·AAL·FAL은 각각 신원확인, 인증, 연합 거래의 보증 강도를 설명하지만, AI 에이전트 거래의 정당성을 모두 설명하지는 못한다. 별도로 위임 범위와 철회 가능성, 에이전트 키와 실행환경의 신뢰성, 이용자 개입 수준, 실행 시점 재검증 여부를 함께 보아야 한다. 예를 들어 두 거래가 모두 높은 AAL을 충족하더라도, 하나는 이용자가 현재 금액과 수취인을 직접 확인한 거래이고, 다른 하나는 일주일 전에 승인한 위임을 바탕으로 에이전트가 무인 실행한 거래일 수 있다. 두 거래의 인증 강도는 같을 수 있지만, 이용자 개입 수준과 실행 위험은 다르다. 따라서 인증 보증 수준만으로 두 실행의 정당성을 동일하게 평가해서는 안 된다.

셋째, 책임 증적은 위험에 비례하면서 프라이버시를 침해하지 않도록 설계해야 한다. 모든 저위험 API 호출에 강한 암호학적 증적을 요구할 필요는 없다. 반면 고액 금융거래, 법률행위, 민감정보 제공처럼 분쟁 가능성과 피해 규모가 큰 거래에서는 승인 조건, 실행 권한, 처리 결과가 서로 검증 가능하게 연결되어야 한다. 중요한 것은 모든 활동을 장기간 추적하는 것이 아니라, 분쟁이나 감사가 필요한 경우 권한 있는 주체가 최소한의 정보로 책임사슬을 재구성할 수 있도록 하는 것이다. 이를 위해 거래별 식별자, 해시 기반 증적, 보존기간 제한, 선택적 증거 제출, 접근권한 통제를 함께 설계해야 한다.

결국 핵심은 AI 에이전트에게 이용자의 접근매체를 넘겨주는 것이 아니라, 이용자가 승인한 범위 안에서만 작동하는 별도의 실행 권한을 만들고, 그 권한의 행사와 결과를 검증 가능한 책임사슬로 남기는 데 있다. 이 방향으로 설계해야 전자금융거래법상 접근매체 규율과 충돌을 줄이면서도, AI 에이전트 기반 거래의 자동화와 책임성을 함께 확보할 수 있을 것으로 보인다.

VI 결론

디지털 신원 지갑은 이용자의 신원과 자격을 증명하는 기술에서 출발했지만, AI 에이전트가 이용자를 대신해 거래를 반복 실행하는 환경에서는 신원 확인만으로 충분하지 않다. 이제는 누가 어떤 권한을 위임했는지, 어느 에이전트가 그 권한을 행사했는지, 실행 결과가 승인된 범위 안에 있었는지를 함께 설명할 수 있어야 한다.

본고는 이를 위해 NIST 가입자 통제형 지갑, EUDI Wallet의 거래별 승인, AP2의 HNP 실행을 비교하고, 신원·권한·의도·에이전트·실행의 다섯 검증축과 신원·자격 확인 → 위임 승인 → 실행 인가 → 실행·재검증 → 결과 증적의 책임사슬을 제시하였다. 또한 RAR를 기반으로 정규화 인가 프로파일과 NAO를 제안하였다.

이 구조에서 사전 승인은 이후의 모든 실행을 자동으로 정당화하지 않는다. 사전 승인은 제한된 실행 권한을 만들기 위한 기준점일 뿐이며, 각 실행 시점에는 NAO, 인가 상태, 실행 시점 문맥, 현재 요청, 에이전트 키 보유 여부를 다시 평가해야 한다. 국내적으로도 이용자의 접근매체 자체를 에이전트에게 맡기는 것보다, 이용자 인증수단과 분리된 에이전트 전용 키와 제한된 실행 권한을 부여하는 구조가 바람직하다.

결국 AI 에이전트 시대의 디지털 신원 지갑은 단순히 '나를 증명하는 지갑'을 넘어 '내가 맡긴 권한을 통제하고, 그 권한으로 AI가 한 일을 설명할 수 있는 지갑'으로 발전해야 한다. 핵심은 이용자의 승인에서 제한된 실행 권한, 실행 시점의 재검증, 결과 증적까지를 하나의 검증 가능한 책임사슬로 연결하는 데 있다.

※ 이 원고는 2026년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.RS-2023-00229400, 안전한 메타버스 환경을 위한 사용자 인증 및 프라이버시 보호 기술개발)

참고문헌

[1] NIST, Digital Identity Guidelines, NIST SP 800-63-4, 2025.
[2] NIST, Digital Identity Guidelines: Federation and Assertions, NIST SP 800-63C-4, 2025.
[3] European Parliament and Council, Regulation (EU) 2024/1183 of 11 April 2024 amending Regulation (EU) No 910/2014 as regards establishing the European Digital Identity Framework, 2024.
[4] European Digital Identity Cooperation Group, European Digital Identity Wallet Architecture and Reference Framework, Version 3.0.0, July 21, 2026.
[5] Google, Agent Payments Protocol (AP2): Agentic Payment Protocol v0.2 Specification, Agent Authorization Framework, Security and Privacy Considerations, AP2 Documentation, accessed Aug. 29, 2026.
[6] T. Lodderstedt, J. Richer, and B. Campbell, OAuth 2.0 Rich Authorization Requests, IETF RFC 9396, May 2023.
[7] D. Fett et al., OAuth 2.0 Demonstrating Proof of Possession (DPoP), IETF RFC 9449, Sep. 2023.
[8] V. Bertocci and B. Campbell, OAuth 2.0 Step Up Authentication Challenge Protocol, IETF RFC 9470, Sep. 2023.
[9] J. Richer and F. Imbault, Grant Negotiation and Authorization Protocol (GNAP), IETF RFC 9635, Oct. 2024.
[10] NIST NCCoE, Accelerating the Adoption of Software and Artificial Intelligence Agent Identity and Authorization, Concept Paper, Initial Public Draft, Feb. 5, 2026.
[11] Marc Anderson and Karën Fort, "Human Where? A New Scale Defining Human Involvement in Technology Communities from an Ethical Standpoint," The International Review of Information Ethics, vol. 31, no. 1, 2022.
[12] 대한민국, 「전자금융거래법」 제2조 제10호, 제6조 및 제9조.
[13] 대법원 2023. 4. 13. 선고 2021도6965 판결.
[14] 대한민국, 「전자금융거래법 시행령」 제8조.