RECON

피싱, 스캠 예방을 위한 서비스 MVP 제안서

BEANNOODLE

LIMITEDEDITION

요약

◆ 팀명
한정판콩국수

◆ 팀원
구본찬, 김도윤, 이현명, 조성민

◆ 프로젝트 이름
RECON

◆ 세부 주제
웹 검색 결과의 다각적 데이터 분석 및 AI 활용을 통해
피싱 여부를 실시간으로 판별하는 브라우저 확장 서비스

◆ 한줄 소개
사용자의 클릭이 안전하도록, 사이트 접속 전의 위험을 정찰합니다.

목차

01 문제 정의

피싱이란

개인정보(Private Data)를 낚는다(Fishing)라는 의미의 합성어

전화 · 문자 · 메신저 · 가짜사이트 등 전기통신수단을 이용하여 피해자를 기만 · 공갈함으로써

이용자의 개인정보나 금융정보를 빼낸 후 금품을 갈취하는 사기 수법

피싱 사이트란

금융 정보, 개인정보를 빼내기 위해 은행, 공공기관 등의 홈페이지와 유사하게 모방한 가짜 사이트

사기범들은 피싱 사이트를 이용하여 금융거래정보의 입력을 유도

01 문제 정의

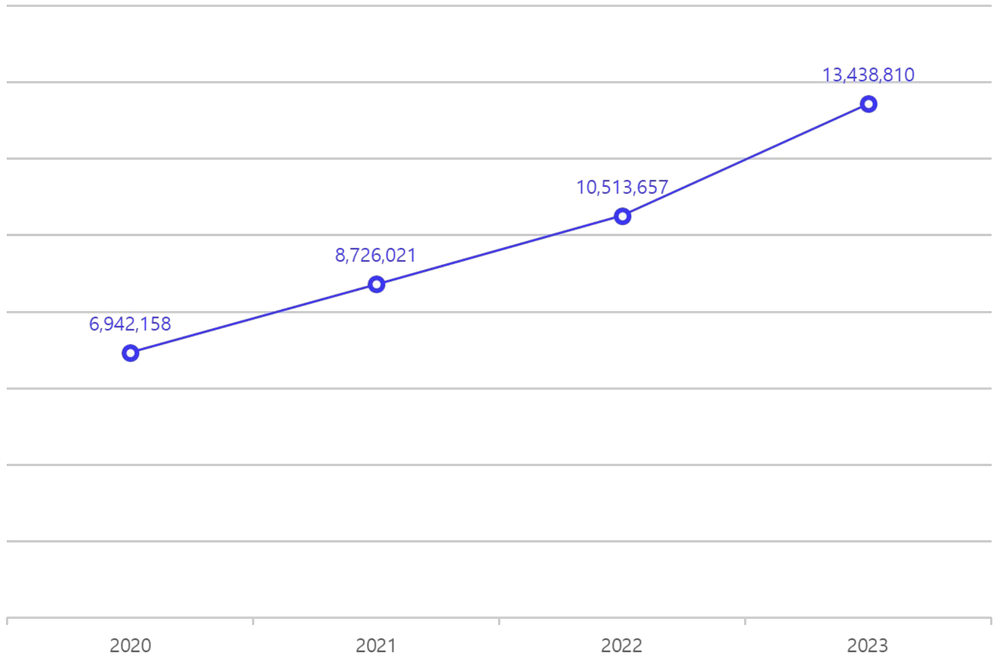
피싱 사이트의 급증 - 해외

Anthropic이 내놓은 보고서에 따르면 ChatGPT가
출시된 2022년 11월 이후 2024년까지 피싱 공격이 4,151% 증가

생성형 인공지능의 등장

↓

웹 페이지 제작 난이도 감소



▲ 사이버 보안 플랫폼 Bolster에서 나온 2024 State of Phishing & Online Scams 보고서에 따르면 글로벌 피싱 페이지가 매해 큰 폭 증가

피싱 사이트와 그 위험도는 매년 증가 추세. 사이버 보안 플랫폼
인 Bolster의 2024 State of Phishing & Online Scams란
보고서에 따르면 글로벌 피싱 페이지가 매해 큰 폭으로 증가

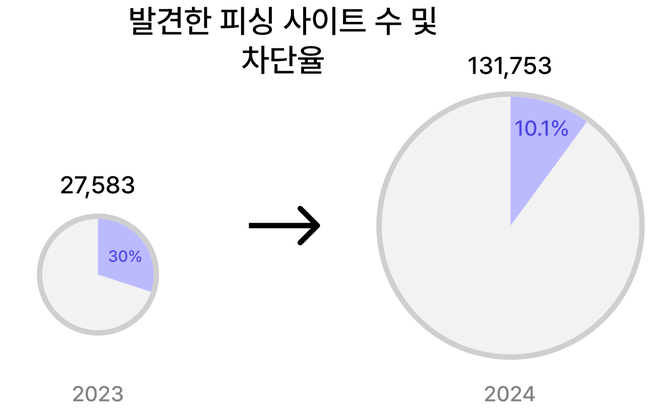
01 문제 정의

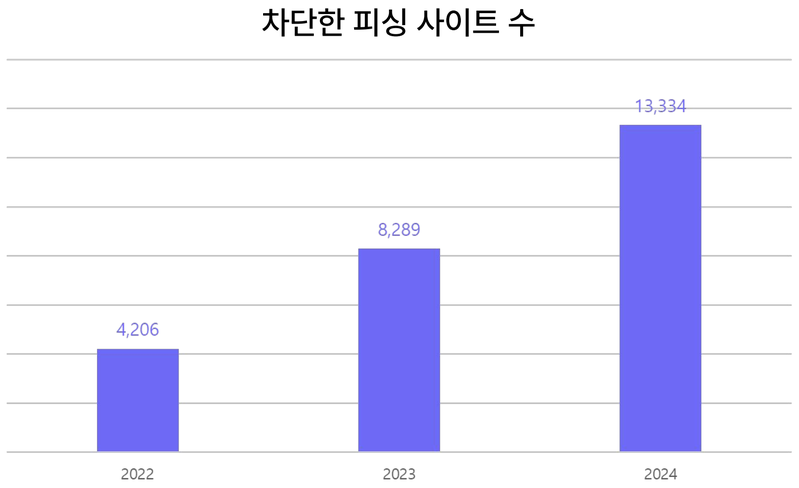
피싱 사이트의 급증 - 국내

KISA에 따르면 발견된 피싱 사이트의 수는 2023년 27,583개에서 2024년 131,753개로 478% 급증

KISA에서 차단한 피싱 사이트의 수 또한 2022년 4,206개, 2023년 8,289개, 2024년 13,324개로 큰 폭으로 증가

그러나 차단율은 2023년 30%에서 10.1%로 크게 하락





01 문제 정의

피싱 사이트의 위험성

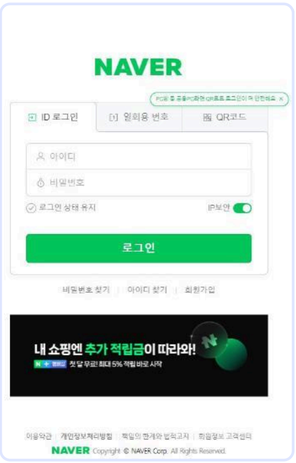
판별 어려움: 피싱 사이트는 그 자체로 일반인의 눈에는 구별이 힘들고 여러 번 Redirecting이 이루어지는 경우도 있어 사이트에 들어가기 전에는 구별하기 어려움

피해 연령대: 모든 연령대에서 피싱 사이트로 인한 피해가 발생하지만, 특히 디지털 보안 이해도가 낮고 금융·서비스 이용이 잦은 고령층과 중장년층의 피해 비중이 크게 나타남

젊은 성인들 또한 많은 활동량으로 인해 피해 비중이 적지 않게 나타남

피싱 사이트

정상 사이트



크리덴셜 스터핑

크리덴셜 스터핑은 유출된 ID와 비번을 이용하여 다른 서비스의 계정을 탈취하는 사이버 공격으로, 사용자의 ID와 비밀번호 유출이 주로 이루어지는 피싱 사이트와 함께 사용되어 그 피해가 극대화됨

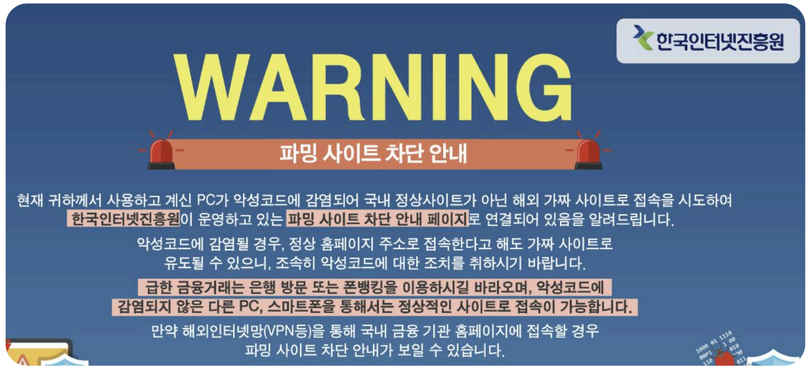
01 문제 정의

KISA의 파밍알리미

악성 코드로 인해 가짜 사이트로 이동된 경우 이를 탐지·안내하는 서비스

성과: 정상 사이트와 가짜 사이트를 구별하기 어렵다는 문제를 해결하여 파밍 범죄를 80% 이상 감소시킴

결론: 피싱·파밍 피해의 주요 원인이 사용자가 위험 사이트를 인지하지 못하는 데 있음을 보여줌



▲ KISA의 파밍 사이트 차단 안내

본 서비스는 이러한 선행 사례를 바탕으로,
사용자가 사이트에 접근하기 전 위험 여부를 사전에 인지할 수 있도록 지원함으로써
피싱 사이트 접속으로 인한 피해를 유의미하게 감소시키는 것을 목표로 함

- 1차 KPI: 사용자 기준 피싱 사이트 접속 피해 건수 80% 이상 감소
- 2차 KPI: 위험 여부 분석 및 안내까지 평균 소요 시간 5초 이내

SOURCE: https://www.boannews.com/media/view.asp?idx=62684

02 제안 솔루션 개요

서비스 이름

RECON

서비스 설명

사용자의 웹 서핑 과정에서 URL 및 웹 페이지 정보를 수집하여 AI 기반으로 피싱 위험도를 실시간 분석하는 크롬 확장 프로그램 기반 보안 서비스

서비스 구성

사용자용 크롬 확장 프로그램, 백엔드 서버, 데이터베이스, 관리자 대시보드, LLM 기반 피싱 판별 모델로 구성

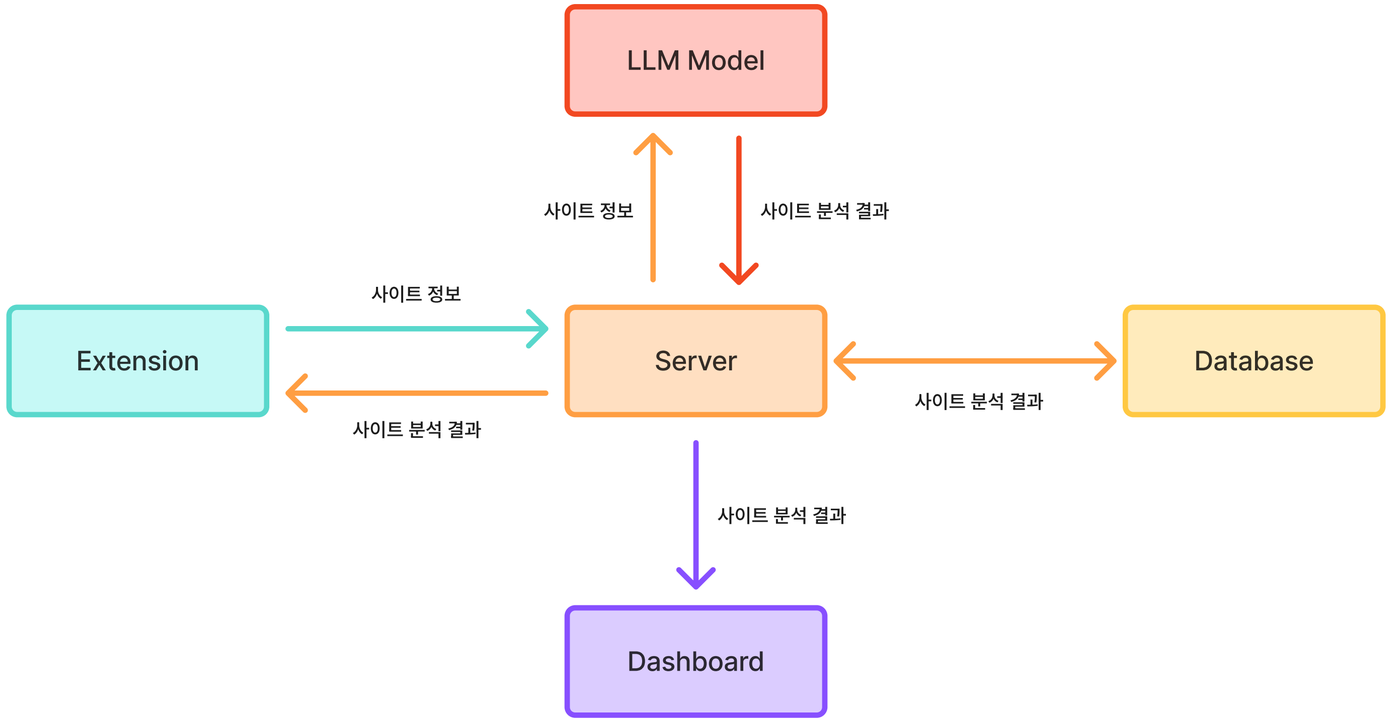
02 제안 솔루션 개요

서비스 구조

크롬 확장 프로그램	사용자와 접촉하는 사이트들의 데이터를 서버로 전송, 받은 분석 결과 표시
백엔드 서버	크롬 확장 프로그램으로부터 받은 데이터를 피싱 판별 모델에 전달 피싱 판별 모델로부터 받은 분석 결과 확장 프로그램에게 전송 분석 결과를 데이터베이스에 저장 관리자 대시보드에게 데이터베이스에 저장된 내용 제공
피싱 판별 모델	서버에게 받은 데이터를 바탕으로 분석 진행 후, 서버에게 전달
관리자 대시보드	서버로부터 받은 데이터베이스에 저장된 분석 내용 기반으로 피싱 사이트 현황, 위험도 통계, 상세분석 등을제공 하여 서비스 이용과 운영 지원
데이터베이스	서버로부터 받은 분석 결과 저장 저장 내용 서버에게 전달 분석 결과는 피싱 사이트 위험도와 판별 사유로 구성되며 축적된 데이터는 향후 모델 고도와 및 관계 기관과의 협업에 활용

02 제안 솔루션 개요

서비스 구조



03 주요 기능 정의

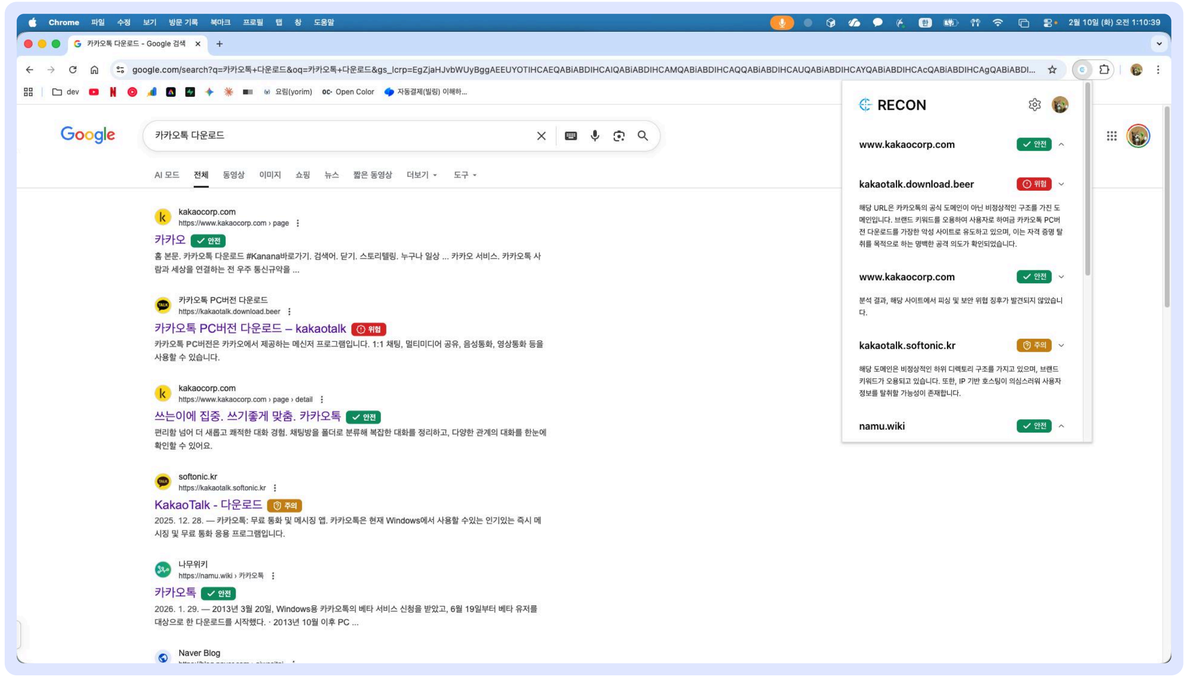
1. 실시간 사이트 탐지 및 위험도 표시

크롬 확장 프로그램을 통해 검색한 사이트들의 데이터를 자동으로 감지

감지한 데이터를 별도의 사용자 데이터 분석 없이 서버로 전송

서버는 전달된 데이터를 기반으로 AI 분석 수행, 피싱 위험도 계산

크롬 확장 프로그램에 피싱 위험도 전달 및 실시간 표시



03 주요 기능 정의

1. 실시간 사이트 탐지 및 위험도 표시

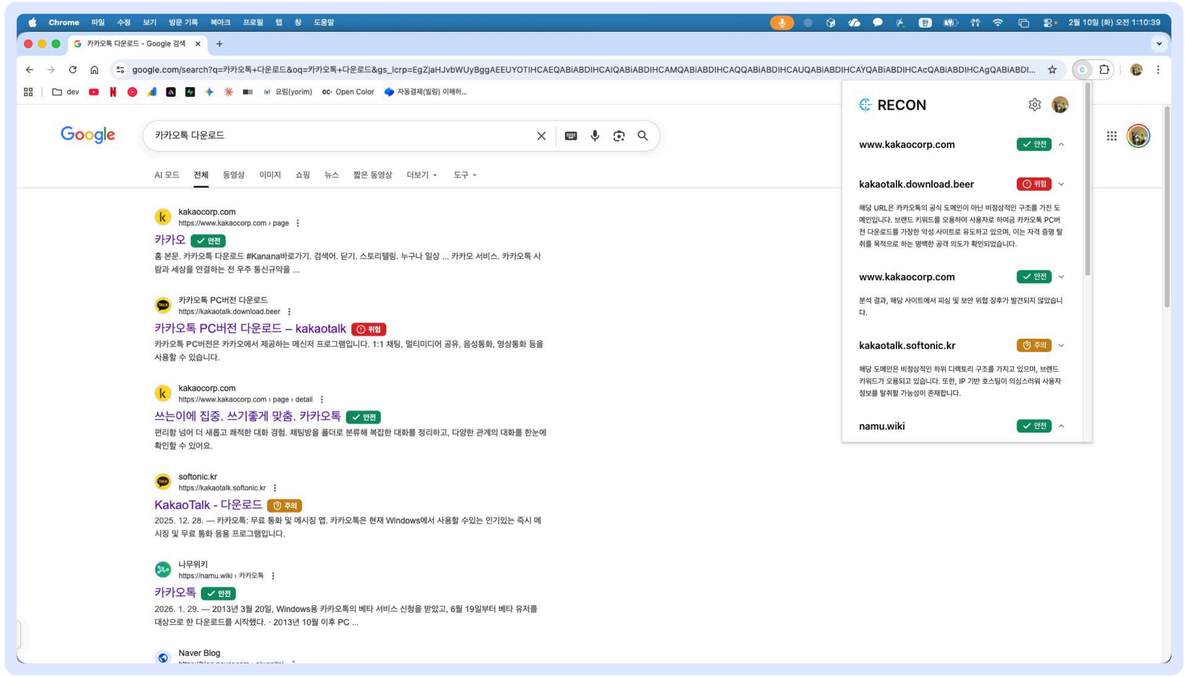
◆ 효과

사용자가 직접 피싱 여부를 판단해야하는 부담 감소

실시간 위험 경고를 통해 무의식적인 피싱 사이트 접속을 사전 차단

◆ MVP 구현 기능

구글 검색 결과 화면에서의 실시간 사이트 탐지 및 위험도 표시



03 주요 기능 정의

2. AI 기반 피싱 판별

- 수집된 URL 및 메타데이터를 기반으로 LLM 모델이 사이트의 피싱 가능성을 분석
- 분석 과정에서는 도메인 정보, URL 구조, 비정상적인 키워드 패턴 등의 판별 사유를 검토하고 이를 바탕으로 피싱 가능성을 산출

◆ 효과

콘텐츠 분석 기반 판별을 통해

새로운 방식의 피싱 사이트에도 대응 가능

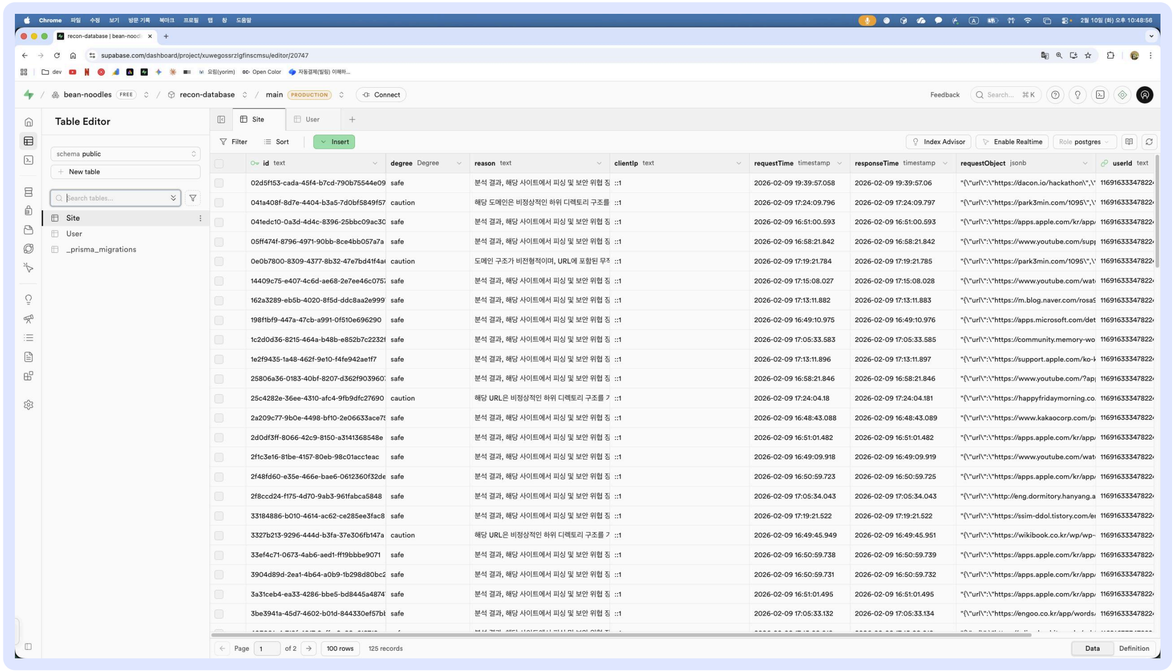
◆ MVP 구현 기능

LLM 기반 피싱 판별 모델을 통한 사이트 분석

03 주요 기능 정의

3. 판별 데이터 저장 및 지속적인 모델 개선

- AI가 측정한 피싱 판별 결과를 데이터베이스에 저장
- 저장 항목은 URL, 제목, 설명 예측된 피싱 위험도, 판별 사유
- 축적된 데이터는 기존 LLM 기반 모델의 성능 개선과
  향후 독립적인 피싱 탐지 모델 개발에 활용
- 피싱 고위험 사이트 정보는 관계 기관과의 협업을 통해
  공유 가능하도록 설계



효과

지속적인 데이터 축적과 학습을 통해 모델 정확도 향상

자체 모델 개발로 외부 의존성을 낮추고 비용 절감, 분석 속도 향상

동시에 달성

MVP 구현 기능

판별 데이터 저장 및 관리자 대시보드

04 데이터 및 기술 활용

AI 모델 학습을 위해 공개 데이터셋 기반의 데이터를 활용

피싱 사이트

정상 사이트

PhishTank 제공 피싱 URL 목록

Tranco Top Sites 제공 상위 웹사이트 URL 목록

메타데이터 추출 및 JSON 변환

실시간 웹 검색이나 외부 평판 정보 없이 피싱 여부를 판단할 수 있도록 데이터 처리 과정을 설계

{ URL, Title, Description }

04 데이터 및 기술 활용

대규모 언어 모델(LLM) 기반 파인튜닝된 분류 모델 사용

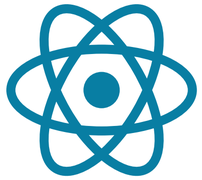
일반적인 머신러닝 기반 피싱 탐지 모델은 독립적인 모델 학습을 위해 대규모 고품질 데이터가 필요하다는 기술적 제약이 존재

문제를 해결하기 위해 LLM 기반 파인튜닝 방식을 활용하여 상대적으로 적은 데이터로도 높은 일반화 성능을 확보

판별된 분석 결과와 데이터를 지속적으로 축적함으로써 독립적인 전용 모델로 확장 가능한 구조를 갖추도록 설계

04 데이터 및 기술 활용

Frontend



React

TypeScript

Vite

Tailwind CSS

Turborepo

Backend

NestJS

Prisma

PostgerSQL

OpenAI

05 사용자 시나리오 / 유즈케이스

주 사용자

온라인 활동이 잦은 일반 웹 사용자

보조 사용자

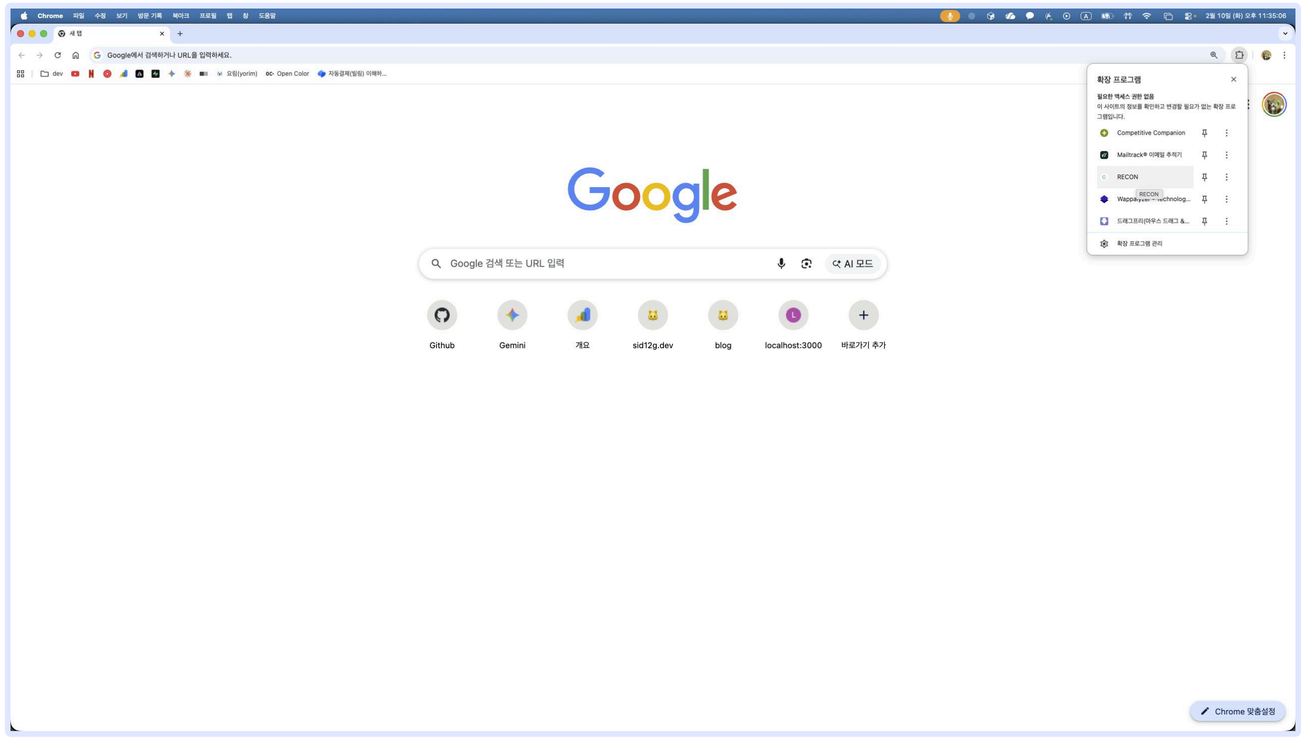
디지털 보안 인식이 낮아 피싱에 취약한 중장년 · 고령층 사용자

사용자 행동 흐름

1. 크롬 확장 프로그램을 설치 후 클릭하여 활성화
2. Google OAuth를 통해 크롬 확장 프로그램 로그인
3. 크롬 브라우저를 통해 웹 서핑을 진행
4. 검색하거나 방문한 사이트 옆에 표시되는 피싱 위험도 뱃지를 확인
5. 의심되는 사이트의 경우 확장 프로그램 팝업을 통해 상세 판별 사유를 확인
6. 위험도가 높은 사이트는 접속을 피하거나 중단하고 주의 단계의 사이트는 개인정보 입력을 자제하는 등 안전한 이용 습관을 유지
7. 실시간 감지를 원하지 않는 경우 설정을 통해 변경, 계정 변경은 계정 관리 페이지를 통해 변경

05 사용자 시나리오 / 유즈케이스

사용자 행동 흐름

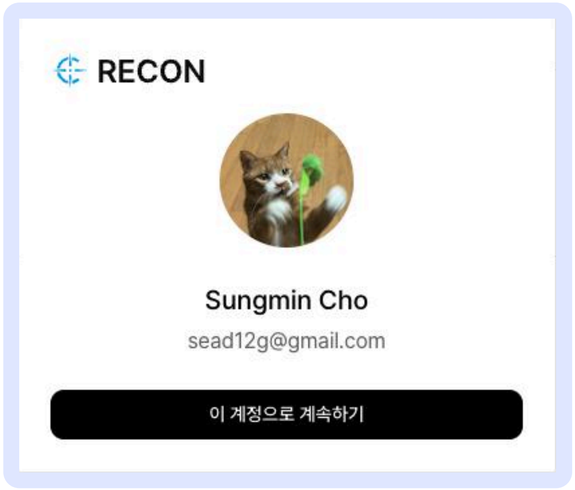


1. 익스텐션 클릭 후 활성화

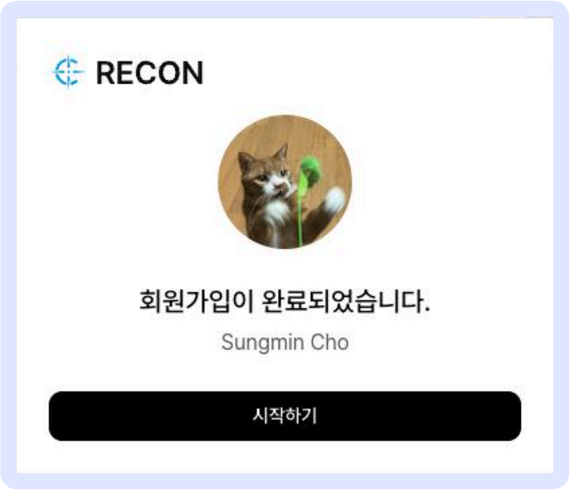
05 사용자 시나리오 / 유즈케이스

사용자 행동 흐름

2. 익스텐션 설치 후 로그인



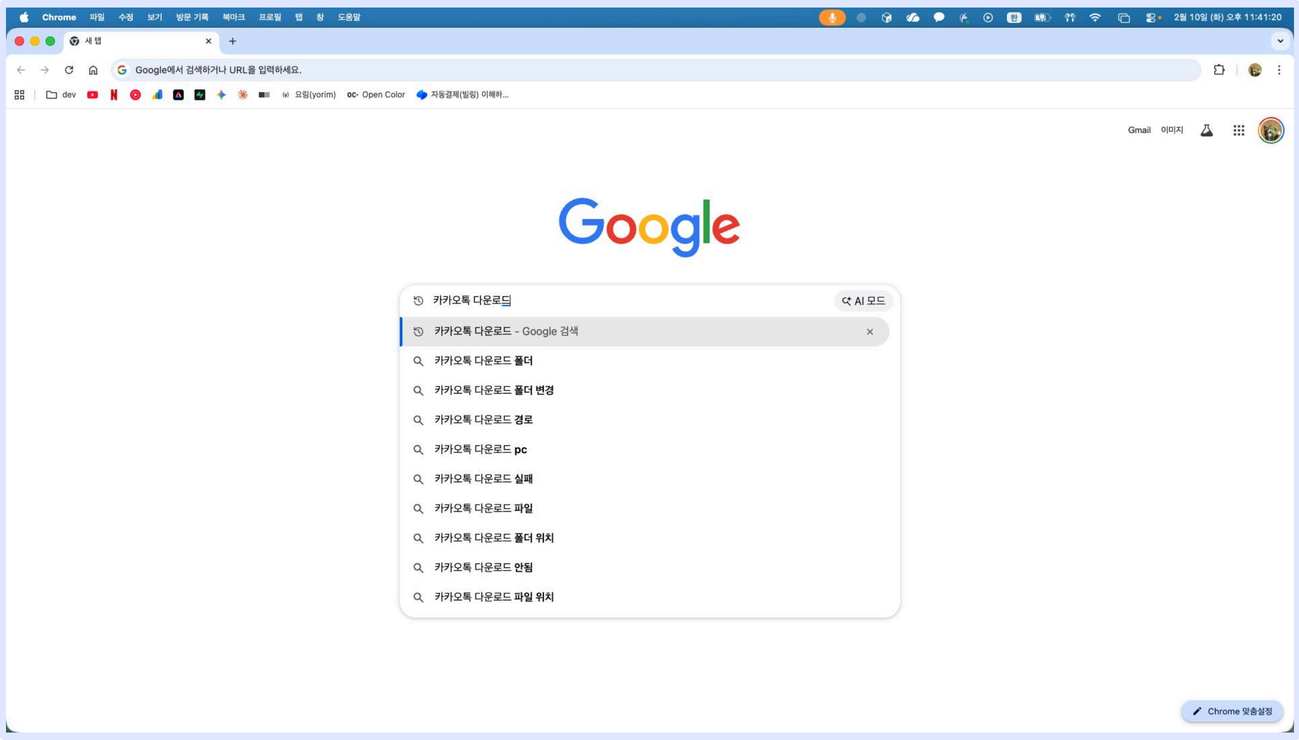
2-a. 계정이 있는 경우 계속하기 클릭



2-b. 계정이 없는 경우 회원가입

05 사용자 시나리오 / 유즈케이스

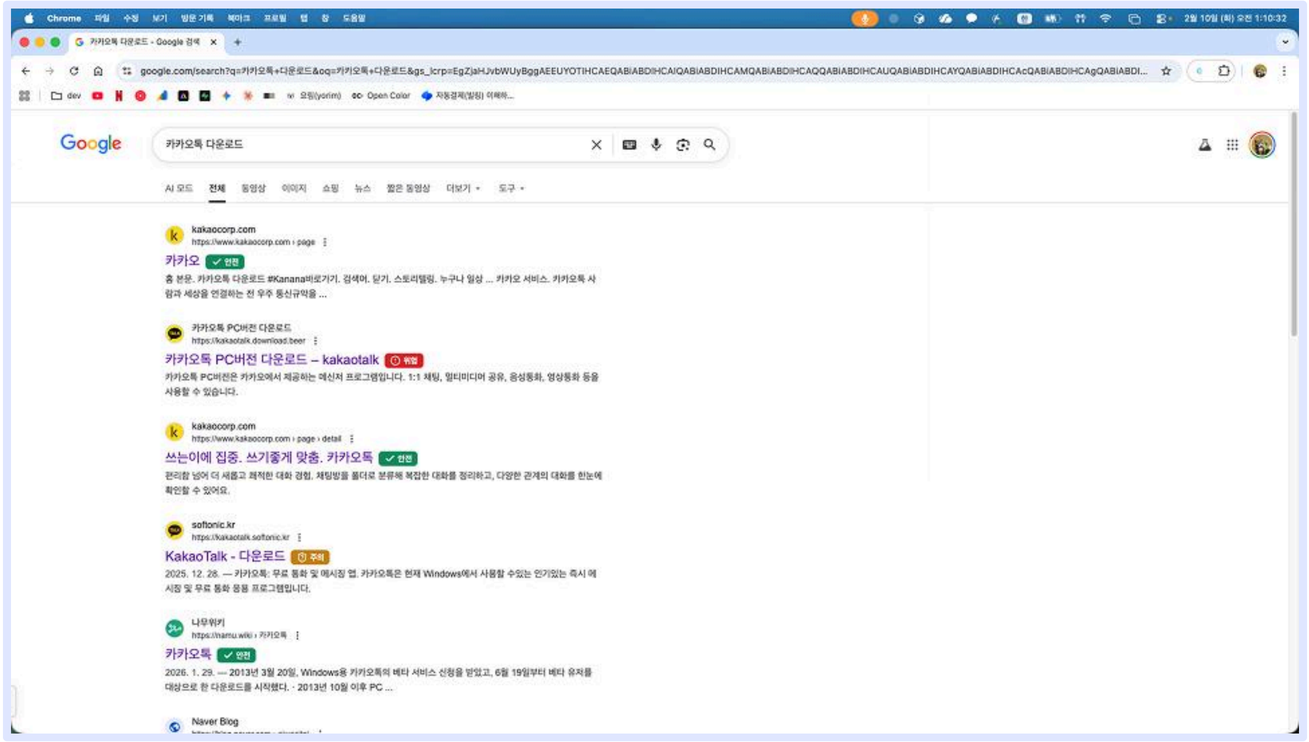
사용자 행동 흐름



3. 크롬 브라우저를 통한 웹 서핑

05 사용자 시나리오 / 유즈케이스

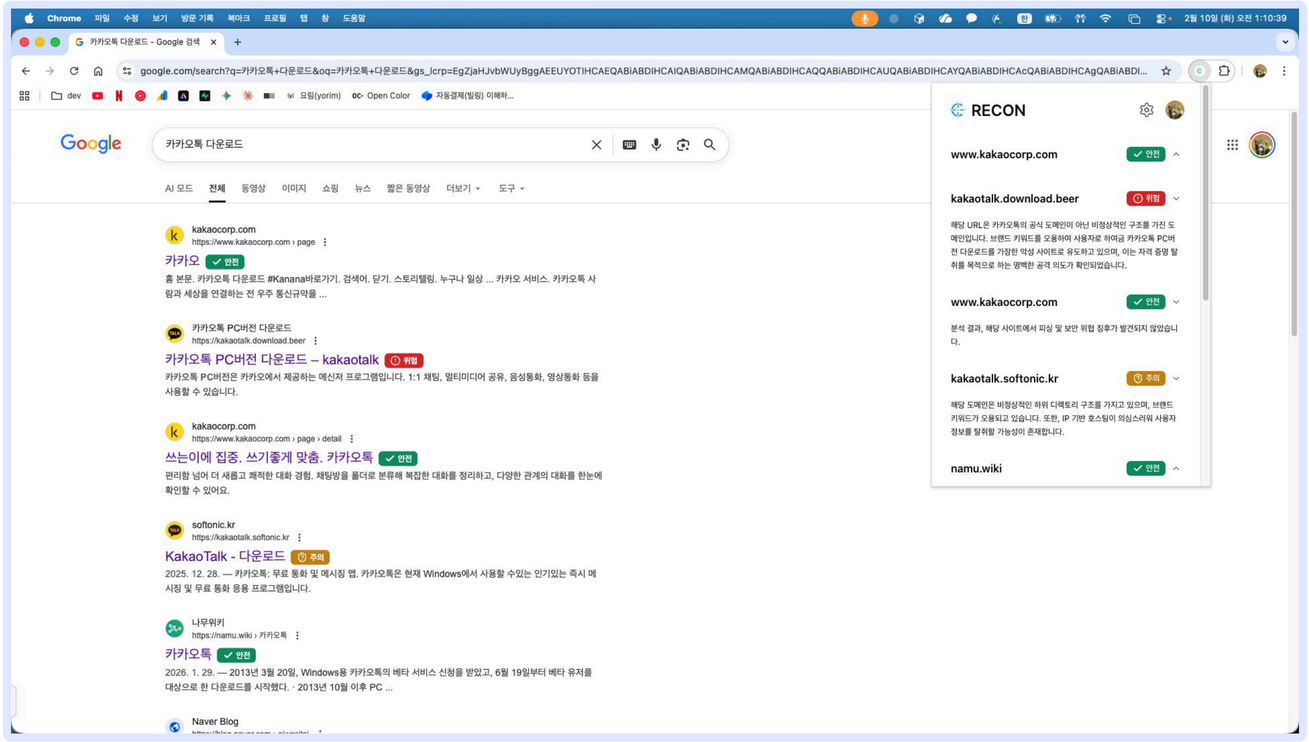
사용자 행동 흐름



4. 피싱 위험도 뱃지 확인

05 사용자 시나리오 / 유즈케이스

사용자 행동 흐름

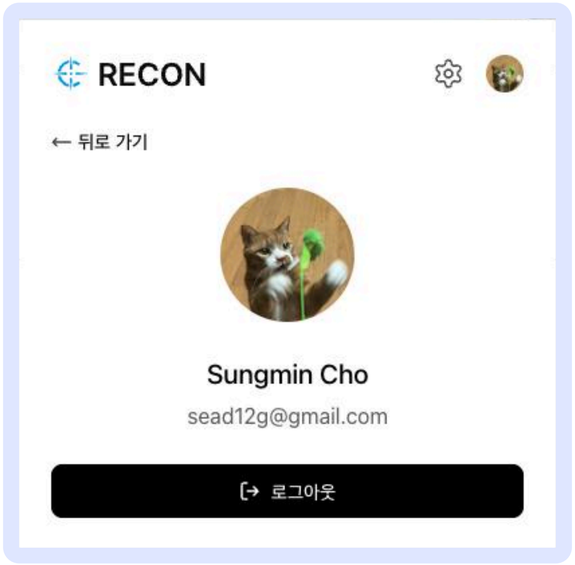


5. 팝업을 통해 상세 판별 사유를 확인

05 사용자 시나리오 / 유즈케이스

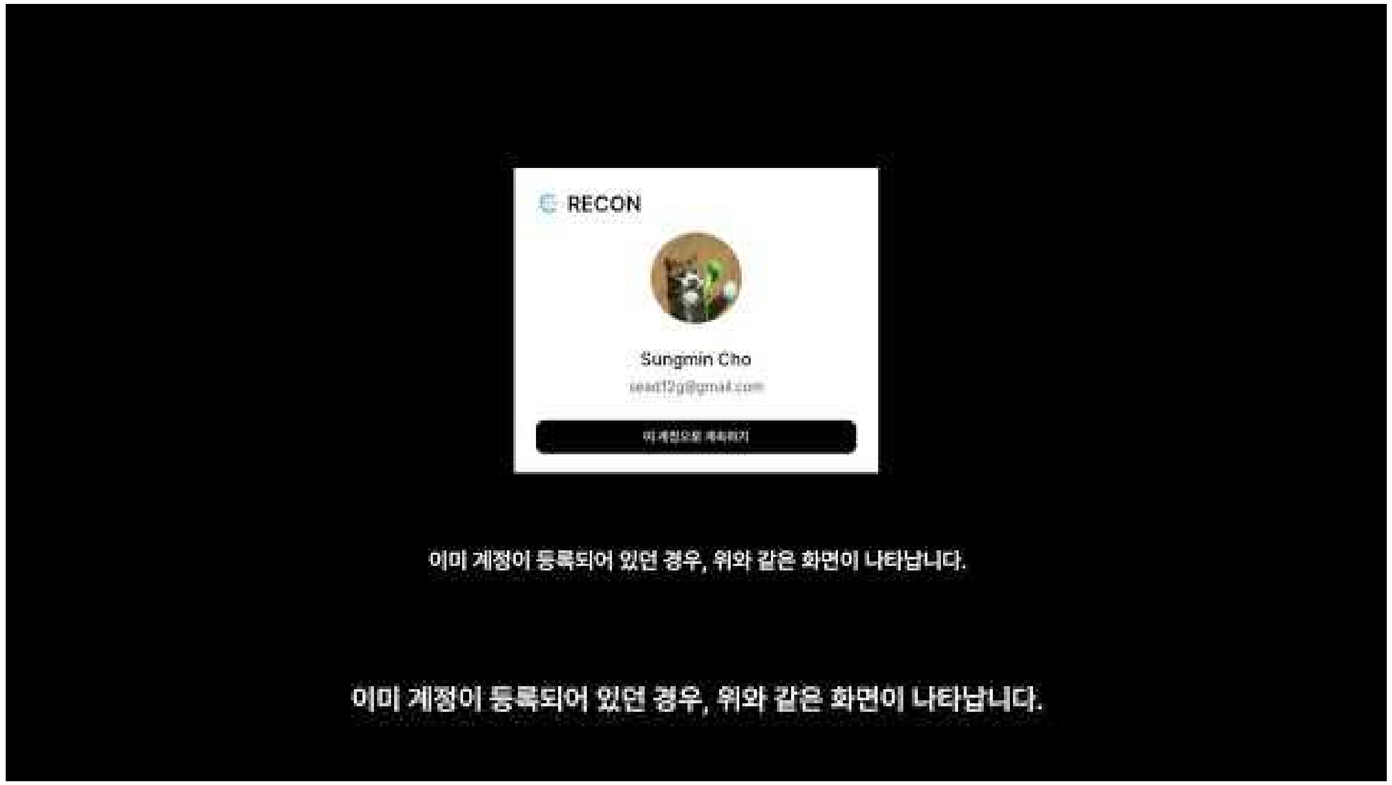
사용자 행동 흐름

7-a. 설정 화면을 통해 검사 여부를 설정



7-b. 계정 관리 페이지에서 로그아웃 가능

05 사용자 시나리오 / 유즈케이스



유튜브 시연 영상

https://www.youtube.com/watch?v=wcKK8alwlIE

06 기대 효과 및 향후 확장성

기대 효과

피싱 피해 건수 80% 이상 감소

피싱 탐지 시간 평균 5초 이내

웹 서핑 과정에서 발생하는 실제 금전 및 개인정보 유출 피해를 대폭 줄일 수 있을 것으로 기대

사용자가 사이트에 접속하기 전 단계에서 위험을 인지하고 사전 대응이 가능

사용자 관점

별도의 보안 지식 없이 실시간 위험도 표시와 판별 사유 제공을 통해 안전한 웹 이용 가능

반복적인 경고 경험을 통해 장기적으로 보안 인식과 이용 습관 개선 효과

사회적 관점

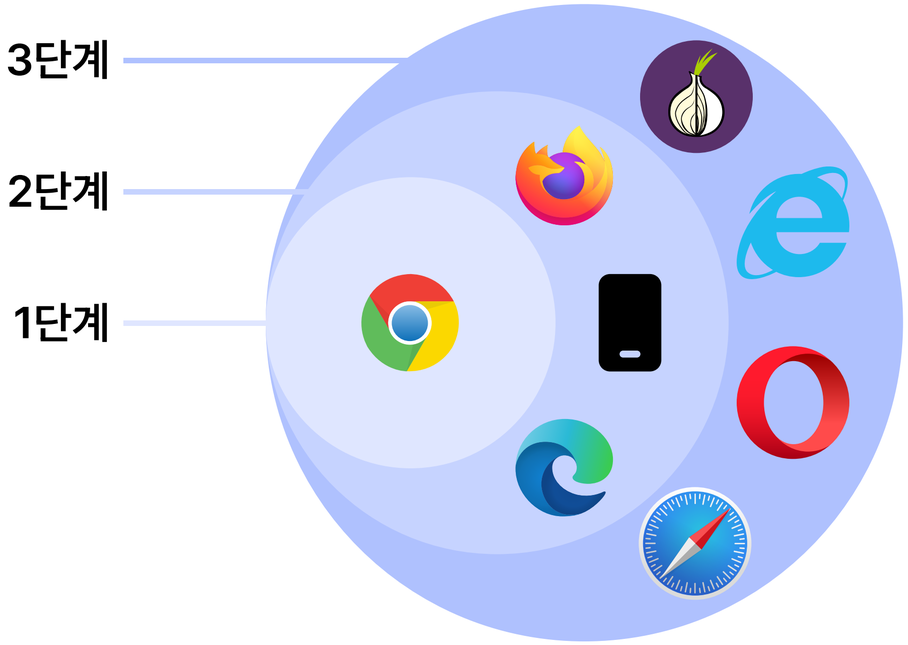
블랙리스트 기반이 아닌 각 사이트의 실시간 데이터 분석을 기반으로 신규 피싱 사이트에 대응 가능

기존의 사후 차단 중심 방식과 달리 사전 탐지 중심의 새로운 피싱 대응 체계 구축 가능

06 기대 효과 및 향후 확장성

확장 가능성

- 초기에는 크롬 확장 프로그램 기반 서비스로 시작하되 향후 Edge, Firefox 등 타 브라우저 확장과 모바일 환경으로 확장
- 경찰청 · KISA 등 공공기관과 연계하여 실시간 피싱 정보 공유에 활용할 수 있으며, 이를 통해 공공 보안 인프라를 지원하는 B2G 보안 솔루션 기업으로의 확장



06 기대 효과 및 향후 확장성

향후 고도화 가능성

초기에는 LLM 기반 판별 모델을
활용하되 판별 데이터를 저장 및 축적

축적된 판별 데이터를 기반으로
독립적인 피싱 판별 모델 개발

Agent 기반 구조를 통해 단순 URL 및
메타데이터 분석을 넘어 페이지 내부
콘텐츠를 사전 탐색

감사합니다

RECON