

## 2025년 새싹 해커톤(SeSAC Hackathon) AI 서비스 기획서

팀명	나노겐
팀 구성원 성명	김현준, 김호영

### 1. AI 서비스 명칭

약관을 읽어주세요

### 2. 활용 인공지능 학습용 데이터

	활용 데이터명	분야	출처
1	TOS_dataset	정책, NLP	<a href="https://huggingface.co/datasets/CodeHima/TOS_Dataset">https://huggingface.co/datasets/CodeHima/TOS_Dataset</a>
2	Lex-Glue - Unfair TOS dataset	정책	<a href="https://huggingface.co/datasets/coastalcp/lex_glue">https://huggingface.co/datasets/coastalcp/lex_glue</a>

### 3. 핵심내용

- 자연어 처리(NLP) 기반의 경량 텍스트 분류 모델을 활용하여 기업 이용 약관의 개별 조항에서 '기업이 사용자에게 악의적으로 손해를 입힐 수 있는 악용 가능성'이 있는 취약점을 자동으로 판별하고 사용자에게 직관적으로 경고하는 서비스 아이디어.

## 4. 제안배경 및 목적

오늘날 개인정보 보호와 관련된 법적 문서, 특히 이용 동의서는 내용이 매우 복잡하고 방대해 일반 이용자가 쉽게 이해하기 어려운 실정입니다. Cranor 와 McDonald 의 연구에 따르면(<https://lorrie.cranor.org/pubs/readingPolicyCost-authorDraft.pdf>), 미국 평균 인터넷 사용자가 1년에 방문하는 약 1,462 개 웹사이트의 개인정보 처리방침을 모두 읽으려면 약 244 시간이 소요되며, 이는 근무일 기준 약 30.5 일, 하루 8 시간이 아닌 실제 인터넷 사용 시간 등을 고려한 추정을 적용하면 약 76 일 정도를 투자해야 하는 수준으로, 현실적으로 모든 정책을 읽는 것이 사실상 불가능한 부담임을 보여줍니다. 이 논문은 이러한 '읽을 수 없는 정책 체계'로 인해 개인정보 처리방침이 실질적인 정보 제공 수단이 아니라, 규제기관이 기대하는 자기결정권 보장을 제대로 수행하지 못하고 있음을 지적하며, 이용자가 자신의 프라이버시를 지키기 위해 마치 금전 대신 '시간'을 지불하고 있다는 점을 사회적 기회비용 관점에서 비판적으로 분석합니다.

국내 조사 결과도 이와 같은 문제를 뒷받침합니다. 한 설문조사에서는 인터넷 이용자의 약 72%가 이용약관을 거의 읽지 않고 동의하며, 68%는 "습관적으로 동의한다"고 응답했고, 72%는 약관 내용을 클릭해서조차 읽지 않는다고 답해, 오늘날 동의 절차가 실질적인 이해를 전제로 한 자기결정 과정이라기보다 형식적인 '체크 박스' 절차에 그치고 있음을 보여줍니다. 이처럼 국민 다수가 약관의 핵심 내용을 제대로 파악하지 못한 채 '묻지마 동의'를 일상적으로 수행하고 있는 현실에서는, 중요 개인정보 처리 절차나 정보주체의 권리·의무를 정확히 이해하지 못한 상태로 민감한 정보 제공, 제 3 자 제공, 해외 이전 등이 이루어질 수밖에 없으며, 이는 개인을 불필요한 위험에 노출시키는 동시에 개인정보 보호법 등 관련 법규의 실질적 준수를 어렵게 만드는 요인이 됩니다.

실제 사례에서도 이러한 위험성이 구체적으로 드러납니다. 예를 들어 청소년 고민상담 앱 ‘나쁜 기억 지우개’는 약관에 이용자의 고민 글이 통계 데이터로 판매될 수 있다는 내용을 숨겨두고, 이를 데이터 거래 사이트에 매물로 올려 사회적 논란을 촉발했으며, 중국 동영상 앱 ‘콰이(快手)’는 국내 이용자의 영상과 얼굴을 별도의 명시적 동의 없이 광고에 활용하면서도 “이용약관상 허용된다”고 주장해, 이용자가 이해하지 못한 동의 조항이 실질적으로 어떻게 악용될 수 있는지를 보여주었습니다. 또한 어떤 해외 앱은 단말기 고유식별번호, 위치 등 민감정보를 대규모로 수집해 해외 서버로 전송하면서도 약관과 개인정보처리방침을 중국어로만 제공해 한국 이용자가 내용을 사실상 확인할 수 없게 했는데, 이는 형식적 동의 구조가 정보 비대칭을 심화시키고, 이용자의 권리를 실질적으로 침해할 수 있음을 잘 보여주는 사례입니다.

크레이너·맥도날드의 연구는, 만약 인터넷 이용자들이 규제기관이 요구하는 이상적인 자기결정 모델을 충실히 따를 경우, 미국 전체 기준으로 연간 약 54 억 시간이 개인정보 처리방침을 읽는 데 소요되고, 이를 시간 가치로 환산하면 약 7,810 억 달러에 달하는 막대한 사회적 기회비용이 발생한다는 점을 계산해 보여줍니다. 이는 같은 시기 미국 온라인 광고 시장 규모(약 210 억 달러 정도)와 비교했을 때, 이용자가 자신의 프라이버시를 보호하기 위해 부담해야 하는 ‘시간 비용’이 산업이 얻는 이익 규모를 훨씬 상회할 정도로 불균형적이라는 점을 시사하며, 현재와 같은 자기책임·동의 중심의 자기결정 구조가 실질적인 권리 보장 장치라기보다, 사실상 현실에서 작동하기 어려운 이상적 모델에 가깝다는 비판으로 이어집니다.

이러한 국내외 연구 및 사례를 종합하면, 오늘날의 개인정보 이용 동의 제도는 이용자의 인지적·시간적 한계를 충분히 고려하지 못한 채 복잡하고 방대한 문서를 일괄 동의하도록 요구함으로써, 형식적 절차만 남긴 채 실질적인 이해와 통제는 확보하지 못하고 있는 것으로 평가할 수 있습니다. 따라서 이용자의 권리 보호와 기업의 법적

책임 완화, 그리고 사용자 신뢰 증대를 위해서는, 법률적·기술적 전문 지식이 없는 일반 이용자도 핵심 내용을 쉽게 이해할 수 있도록 개인정보 이용 동의서와 개인정보처리방침을 자동으로 분석·요약하고, 민감한 조항 및 잠재적 위험을 직관적으로 설명해 주는 AI 기반 지원 시스템이 필요합니다. 이러한 AI 시스템은 사용자가 동의서의 실제 의미와 위험도를 명확히 파악한 뒤 스스로 판단할 수 있도록 돕고, '깜깜이 동의' 관행을 줄이며, 개인정보 처리 과정의 투명성과 정보주체의 실질적 권리 행사를 강화함으로써, 디지털 시대에 적합한 신뢰 기반의 개인정보 보호 환경 조성에 기여할 수 있을 것입니다.

## 5. 세부내용

구분	세부 내용
모델 구조 및 목표	자연어 처리(NLP) 기반의 <b>경량 텍스트 분류 모델</b> 을 개발합니다.
	- <b>기반 모델:</b> 학습 및 추론 속도, 그리고 Chrome Extension 환경에서의 경량화를 고려하여 <b>DistilBERT</b> 모델을 선정합니다.
	- <b>법률적 맥락 보존:</b> <a href="#">Legal_Specific_KoLLM</a> 과 같은 법률 특화 모델의 접근 방식을 참고하여 모델의 법률 이해도를 높여 정교한 분석을 수행합니다.
	- <b>모델 목표:</b> 약관의 개별 문장(조항)을 입력받아 '기업의 악용 가능성이 있는 취약 조항', '일반 조항' 등의 다중 클래스(MultiClass)로 분류하는 것을 목표로

	합니다.
학습 데이터	제안 배경에서 명시된 <b>TOS_dataset</b> 및 <b>Lex-Glue - Unfair TOS dataset</b> 을 활용하여 모델을 파인튜닝(Fine-tuning)합니다.
모델 최적화 및 배포 환경	<ul style="list-style-type: none"> <li>- <b>로컬 구동 환경:</b> 최종 모델은 사용자의 로컬 환경(Local)에서 구동되도록 설계하여 서버 부하 및 외부 전송 없이 개인정보 보호 관련 조항을 즉각적으로 분석합니다.</li> <li>- <b>최적화:</b> 모델 배포 시 <b>ONNX</b> 형식으로 변환하고, 추론 속도 향상 및 용량 최소화를 위해 양자화(Quantization)를 필수적으로 적용합니다.</li> </ul>
서비스 제공 방식	<ul style="list-style-type: none"> <li>- <b>플랫폼:</b> 개발 및 최적화된 경량 모델을 탑재한 <b>Chrome Extension(크롬 확장 프로그램)</b> 형태로 사용자에게 배포합니다.</li> <li>- <b>기능:</b> 사용자가 접속한 웹페이지의 이용 약관 및 개인정보 처리방침 텍스트를 입력받아 실시간으로 분석합니다.</li> <li>- <b>시각화:</b> 모델 분석 결과, '부당한 어구' 또는 '악용 가능성이 있는 취약 조항'으로 분류된 텍스트에 대해 사용자에게 직관적인 시각화(예: 하이라이팅, 경고 아이콘 표시)를 제공하여, 사용자가 '묻지마 동의' 대신 위험성을 인지하고 판단할 수 있도록 지원합니다.</li> </ul>

## 6. 기대효과

### ○ 사회/경제적 파급(기대) 효과

- **사회적 가치:** 일반 사용자의 정보 불평등을 해소하고, 기업의 일방적인 약관 개정이나 불공정 거래 행위로부터 소비자의 권익을 선제적으로 보호하는 데 기여합니다.

- **경제적 가치:** 사용자 친화적인 약관 작성은 기업에 유도하여 투명한 거래 환경을 조성하고, 장기적으로는 약관 관련 분쟁 비용을 감소시키는 효과를 기대할 수 있습니다.

※ 상세 설명을 위해 도표, 스케치 등 별도파일 추가 가능

※ 제출한 기획서는 온라인 예선 심사 전 구체화하여 깃허브(GitHub)에 필수로 게시