JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 29 No. 10, pp. 77-87, October 2024
https://doi.org/10.9708/jksci.2024.29.10.077

A Study on Measuring the Risk of Re-identification of Personal Information in Conversational Text Data using AI

Dong-Hyun Kim*, Ye-Seul Cho**, Tae-Jong Kim***

*Professor, Dept. of AI Information Security, Halla University, Gangwon-do, Korea
**Senior research engineer, World Vertex Co., Ltd, Seoul, Korea
***CEO, World Vertex Co., Ltd, Seoul, Korea

[Abstract]

With the recent advancements in artificial intelligence, various chatbots have emerged, efficiently performing everyday tasks such as hotel bookings, news updates, and legal consultations. Particularly, generative chatbots like ChatGPT are expanding their applicability by generating original content in fields such as education, research, and the arts. However, the training of these AI chatbots requires large volumes of conversational text data, such as customer service records, which has led to privacy infringement cases domestically and internationally due to the use of unrefined data. This study proposes a methodology to quantitatively assess the re-identification risk of personal information contained in conversational text data used for training AI chatbots. To validate the proposed methodology, we conducted a case study using synthetic conversational data and carried out a survey with 220 external experts, confirming the significance of the proposed approach.

▸ **Key words**: Personal Information, AI Chatbot, Conversational Data, Risk Measurement

[요 약]

최근 인공지능 기술 발전으로 다양한 챗봇이 등장하여 호텔 예약, 뉴스 확인, 법률 상담 등 일상 작업을 효율적으로 수행하고 있다. 특히 ChatGPT와 같은 생성형 챗봇은 교육, 연구, 예술 분야에서 자체 콘텐츠를 생성하는 등 활용 가능성을 확장하고 있다. 이러한 AI챗봇의 학습에는 고객 서비스 대화 기록 등 방대한 양의 '대화형 텍스트 데이터'가 필요하지만, 정제되지 않은 대화형 텍스트 데이터의 학습으로 인해 국내외에서 AI챗봇에 대한 개인정보 침해 사례가 발생하고 있다. 본 연구는 AI챗봇 학습에 사용되는 '대화형 텍스트 데이터'를 기반으로 데이터 내 포함되어 있는 개인정보 항목에 대한 재식별 위험성을 계량적으로 측정할 수 있는 방법론을 제안하고 있다. 제안 방법론에 대한 타당성 검증을 위해 가상의 대화형 데이터를 생성하여 자체실증을 하였으며, 외부 전문가 220명을 대상으로 설문조사를 실시하여 제안하는 방법론의 유의미함을 확인할 수 있었다.

▸ **주제어**: 개인정보, AI챗봇, AI학습데이터, 대화형 데이터 위험측정, 재식별 위험 측정

• First Author: Dong-Hyun Kim, Corresponding Author: Dong-Hyun Kim
*Dong-Hyun Kim (dh.kim@halla.ac.kr), Dept. of AI Information Security, Halla University
**Ye-Seul Cho (yesl-c@vtex.co.kr), World Vertex Co., Ltd
***Tae-Jong Kim (ktj4820@gmail.com), World Vertex Co., Ltd
• Received: 2024. 08. 05, Revised: 2024. 10. 07, Accepted: 2024. 10. 07.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

최근 인공지능 기술 발전에 따라 일상적인 작업을 효율적으로 수행할 수 있는 다양한 챗봇(Chatbot)이 등장하고 있다. 이러한 챗봇은 메신저 기반 환경에서 사용자들의 질문이나 요구사항에 대하여 자동으로 응답을 제공해 주는 서비스[1]로써, 초기에는 키워드에 대한 정해진 답변을 제공하는 단순한 형태였으나 최근에는 자연어 처리(NLP : Natural Language Processing), 머신러닝, 심층 학습과 같은 첨단 기술을 활용하여 보다 복잡하고 개인화된 상호작용을 제공하고 있다. 최근 챗봇 시장은 호텔 예약, 뉴스 확인, 법률 상담 등 다양한 분야로 사용 범위가 확대되고 있으며[2], 2023년 개발된 ChatGPT와 같은 고도화된 AI 챗봇으로 인해 교육, 연구, 예술 등의 분야에서 자체적인 콘텐츠를 생성함으로써 AI챗봇의 활용 가능성을 한층 더 확대하였다.

글로벌 시장조사업체인 프레시던스 리서치(Precedence Research)는 Fig. 1과 같이 글로벌 AI챗봇 시장 규모가 2022년 8억 4,000만 달러에서 연평균 19.29% 성장해 2032년에는 49억 달러에 이를 것으로 전망하고 있으며[3], AI챗봇은 단순한 보조 도구를 넘어 필수적인 서비스 제공자로써의 역할을 확립하고 있다.

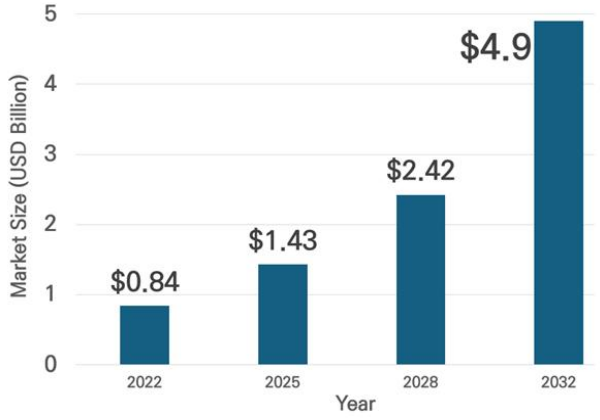

Fig. 1. Trends in AI Chatbot Market Size

AI챗봇의 학습에는 방대한 양의 '대화형 텍스트 데이터' 확보가 필수적이다. 이 데이터는 챗봇이 다양한 맥락에서 적절한 응답을 생성하고, 사용자 의도를 정확하게 이해할 수 있도록 도움을 준다. '대화형 텍스트 데이터'는 주로 고객 서비스의 대화 기록, 소셜 미디어 상의 대화, 온라인 포럼 등의 다양한 소스에서 수집되고 있으며 AI챗봇 학습모델을 훈련시키는데 핵심 요소가 되고 있으나, 최근 AI챗봇을 이용한 다양한 개인정보 유출 및 윤리적 문제 등이 발생하고 있다. 2017년 마이크로소프트의 AI챗봇인 Tay는 트위터 사용자들과 상호 작용을 하면서 혐오 발언을 학습해 문제가 일어났으며[4], 국내에서는 스케터랩의 AI챗봇인 이루다가 이용자들의 성희롱 문제와 인종차별 등 사회적 혐오를 부추긴다는 우려 및 개인정보가 유출되었다는 의혹 등의 문제로 발표 이후 3주 만에 서비스를 중단하였다[5].

이러한 문제점의 원인으로는 첫째, 흑인·동성애자·장애인·여성 등에 관한 질문에 AI챗봇이 혐오 표현에 동조하고 차별적 표현을 한 점과 둘째, 일부 사용자들이 AI챗봇을 대상으로 악의적 대화를 유도한 점, 그리고 마지막으로 셋째, AI챗봇에 제공된 데이터에 개인정보가 포함되어 개인의 프라이버시를 침해했다는 점이다[6]. AI 챗봇의 혐오 표현 문제나 일부 사용자가 AI 챗봇에게 악의적인 대화를 유도하는 문제는 AI 챗봇 개발 시 편향적이거나 악의적인 질문을 필터링하거나 거부하는 메커니즘을 통해 방지할 수 있다. 하지만 세 번째 문제의 경우 사전에 학습을 위해 수집한 데이터에 개인정보가 포함되어 있다는 것을 판단하기는 매우 어려운 문제일 수 있다.

이를 개인정보 활용 관점에서 살펴보면 국내 '개인정보보호법'에 따라 과학적 목적으로 대화형 텍스트 데이터를 가명처리하여 AI학습 등의 목적으로 활용할 수 있다. 하지만 데이터 내에 법률에서 정의하고 있는 식별정보(Identifier)와 식별가능정보(Quasi-Identifier)를 구분하는 것은 매우 어려운 문제이다. 데이터가 활용되는 환경에 따라 개인정보로 판단할 수도 있고, 아닐수도 있기 때문이다. 특히, 대량의 비정형 데이터의 경우 다양한 식별가능정보(준식별자)가 존재하고 있으며, 인력을 통한 전수검사가 현실적으로 불가능하다. 또한, 개인정보의 식별 위험성을 적절히 검토했다 하더라도 비식별 조치 관점에서의 문제도 남아있다. AI학습을 위한 대상 개인정보의 재식별 가능성을 줄이기 위해 비식별 조치 수준을 높이면 보호 측면은 강화되지만, 빅데이터로 활용할 수 있는 데이터의 품질은 떨어지게 된다. 결과적으로 개발한 AI챗봇의 성능이 저하될 수밖에 없는 것이다. 그와 반대로 데이터의 품질을 높이기 위해 비식별 조치 수준을 낮추게 되는 경우 활용하는 데이터에 개인식별가능 정보가 포함될 수 있으며, 향후 활용하는 과정에서 의도치 않은 개인정보나 특정 개인을 식별할 수 있는 '다른 정보[1)]'들이 발화할 수 있다.

1) 가명처리 전 정보(원본)와 추가정보를 제외한 개인정보처리자가 보유하고 있는 정보를 말함(가명정보 처리 가이드라인 p.15)

본 연구는 위와 같은 문제점을 해결하기 위해 AI챗봇 등을 학습시키는데 주연료가 되는 '대화형 텍스트 데이터'를 대상으로 사전에 개인정보가 포함된 요소를 분석 및 측정하여 향후 활용 과정에서 발생할 수 있는 재식별 가능성을 줄이고, 안전하게 '대화형 텍스트 데이터'를 활용하는 것을 목표로 하고 있다. 이를 위해 제2장에서 개인정보를 비식별 조치하여 활용함에 있어 발생할 수 있는 재식별 가능성을 줄이기 위해 고려해야 하는 사항 등을 국내외 정책을 통해 살펴보고, 데이터 자체의 개인정보 위험성을 기술적으로 측정할 수 있는 방안에 대해 살펴보고자 한다. 이어 제3장에서 '데이터 자체'와 '데이터 환경(Context)' 및 '대화형 텍스트 데이터'의 특성을 고려한 재식별 위험도 측정 방법론을 제시하고, 제4장에서는 가상 데이터를 활용한 자체 실증 및 전문가 설문조사를 통해 제안하는 방법론에 대한 타당성과 효과성을 검증하여 제5장을 통해 결론을 제시하고자 한다.

II. Preliminaries

1. Policies related to Re-Identification of Personal Information

1.1 UKAN Anonymisation Decison-Making Framework [7]

영국 ICO(Information Commissioner's Office)는 2012년 '개인정보 익명화 실무지침'을 발간하고, UKAN(UK Anonymisation Network)은 이러한 지침을 구체적으로 설명하고 실무자가 현장에서 활용할 수 있는 내용을 보완하여 2016년 '익명화에 관한 의사결정 프레임워크(The Anonymisation Decison-Making Framework)'를 마련하였다. 본 프레임워크의 주요 특징으로는 효과적인 익명화 처리를 위해서는 데이터가 활용되는 환경을 고려하여야 하며, 이러한 환경에 따라 식별이 가능한 정보와 식별이 불가능한 정보를 구분하여야 한다고 제시하고 있다. 해당 지침에서는 익명화 유형을 형식적(Formal), 보장된(Guaranteed), 통계적(Statistical), 기능적(Functional)으로 4가지 유형을 제시하고 있는데, 마지막으로 제시된 기능적 익명화는 2020년 개인정보 보호법 개정에 따라 마련된 국내 '가명정보 처리 가이드라인[8]'에서 가명 처리 대상 정보의 개인정보 식별 위험성 검토 절차의 기초모델로 활용되었다.

1.2 Guidelines for Processing Pseudonymization Information [8]

해당 가이드라인은 '개인정보 보호법'에 따라 가명정보를 안전하게 처리할 수 있는 절차와 기술 정의를 제공하고 있으며, 금융 · 보건의료 · 교육 등 분야별 가이드라인의 베이스 모델이 되는 지침이다. 해당 지침에서는 개인정보 재식별 위험성을 경감하기 위해 관리적 보호조치로 '사전준비' 단계에서 적법한 활용 범위 및 책임소재(계약서 등) 등을 명확히 규정하도록 하고 있으며, 기술적으로는 '위험성 검토' 단계를 통해 데이터 자체의 식별 위험성과 처리환경의 식별 위험성을 검토하도록 제시하고 있다. 종합적으로 기술적 검토 과정에서 데이터의 양과 분포, Outlier 등을 정량적으로 판단하고 있지만, 최종 검토 의견은 외부 전문가를 통해 정성적으로 판단하고 있어 평가(심의)위원의 전문성에 따라 위험성의 판단기준이 상이하게 적용된다는 단점을 가지고 있다.

1.3 NIST SP800-188 [9]

미국 NIST는 정부 데이터셋을 비식별화하여 활용하기 위한 구체적인 절차 및 평가 방법을 제시하고 있다. 비식별 정보의 위협평가 및 재식별 가능성 검토 등 8단계의 절차에서 개인정보가 재식별 되지 않도록 검토해야 하는 구체적인 사항을 제안하고 있으며, 특히 데이터 공유 모델(Data Sharing Model)을 통해 데이터가 공개되는 유형에 따른 고려사항을 제안하고 있다. 해당 지침의 특징으로는 개인정보 비식별 처리가 재식별 가능성을 0%로 만드는 것이 아니라 고의적인 데이터 공개를 통해 발생할 수 있는 재식별 공격 위협을 최소화(경감)하는데 그 목적이 있다고 제시하고 있다.

1.4 ISO/IEC 25237 [10]

ISO/IEC 25237은 의료정보의 안전한 2차(목적 외) 이용을 위한 가명화(Pseudonymization) 처리 절차 및 방법, 시나리오 등을 제시하고 있는 국제표준이다. 해당 표준의 특징으로는 개인정보 보호의 보증 수준에 따른 가명처리 절차를 3가지 수준으로 제시하고 있는데 Level.1은 데이터 요소에 의해 식별 가능한 대상자와 관련된 위험도로 개인정보의 가장 낮은 수준의 보호를 제시하고 있다. 처리 기준으로는 명백히 식별되는 정보 및 간접적으로 식별되는 정보를 제거하는 것이다. Level.2는 데이터 변수들의 통합된 정보와 관련된 위험도로 외부 정보를 활용(이용 · 연계)하는 공격자에 대한 방어를 고려하고 있다. 해당 레벨에서는 global data model과 내부 정보의 흐름을 고려

하여야 하고, 외부정보(Observational data)가 어떤 특징을 가지고 있는지에 따른 처리 수준을 고려하여야 한다. Level.3의 경우 정보집합물의 이상치와 관련된 위험도로 Outlier나 Rare 데이터를 통한 재식별 가능성을 고려하여 처리하도록 제시하고 있다.

1.5 Other Policies

우리나라의 공공기관은 '개인정보 보호법'에 따라 개인정보처리시스템을 구축하는 경우 영향평가를 받도록 되어 있다. 이러한 영향평가를 원활하게 수행할 수 있도록 '개인정보 영향평가 수행 안내서[11]'를 제시하고 있는데, 해당 안내서에서는 개인정보 항목별 영향도를 등급표로 작성하여 1~3등급으로 구별하여 관리해야 한다. 개인정보 영향평가에서의 위험도 측정은 침해요인의 발생 가능성(Possibility of Infringement Factors)과 법률 준수사항(Legal Compliance)을 주요 위험으로 파악하여 가중치를 부여하고, 처리 업무의 중요도(Persinal Information Impact)를 합산하여 Fig. 2와 같이 계산된다.

$$Risk = Personal\,Information\,Impact + (Possibility\,of\,Infringement\,Factors \times Legal\,Compliance) \times 2$$

Fig. 2. Estimating the risk of PI infringement

그 외 2020년 개인정보보호위원회에서 발간된 '개인정보 위험도 분석 기준 및 해설서[12]'에서는 개인정보처리자가 보유한 고유식별정보 보유 여부 및 법률에서 제시하고 있는 보호조치에 대한 이행 여부만을 검토하여 암호화 적용 여부를 검토하기 위한 기준을 제시하고 있으며, 홈페이지 개인정보 노출방지 안내서[13]에서는 개인정보의 위험성이 아닌 홈페이지를 통한 노출 위험성을 측정하는 방안을 제시하고 있어 일반적인 개인정보에 대한 위험성을 측정하는 기술 내용은 찾아볼 수 없었다.

2. Technology related to Re-Identification of Personal Information

2.1 PDRA(Privacy Detection and Risk Analysis)

Kim [14]은 개인정보 탐지 및 위험분석 모델에 대한 연구를 통해 PDRA모델을 제안하였다. PDRA란 PC에 기록된 개인정보 항목을 탐지하고 그에 따른 개인정보 중요도를 평가하여 사용자에게 개인정보 유출 시 발생할 수 있는 위험 수준을 고지하고, 사용자의 선택에 따라 암호화 등의 안전조치를 통해 개인정보를 안전하게 보호할 수 있음을 정의하고 있다. PDRA는 1. PIDM(Privacy Information Detection Mechanism), 2. PIPM (Privacy Information Patten Analysis & Parsing Mechanism), 3.PIRM(Privacy Information Risk Management Mechanism), 4.CEM(Crypto Enhanced Mechanism) 등 4가지로 구성하고 있다. 주요 알고리즘은 개인정보 탐지의 경우 패턴을 이용하여 파싱하였으며, 개인정보 속성의 민감도에 따른 위험성 측정을 노출 시 금전적인 피해, 개인정보 도용, 사생활 침해 등으로 구분하여 P1~P4의 위험도로 분류하였다. 마지막으로 4단계로 분류한 개인정보 항목의 조합을 통해 최종 위험도를 Level 1~8까지 분류하여 측정하고 있다.

2.2 Personal Information Risk Using IPA

Jeong [15]등은 개인정보의 위험도를 분석하기 위한 도구로 TF-IDF(Term Frequency-Inverse Document Frequency) 기법과 IPA(Importance Performance Analysis) 기법을 활용하였다. TF-IDF 모델은 1972년 Sparck Jones[16]에 의해 제안되었으며 문헌빈도에 용어빈도를 곱하여 단어에 가중치를 부여하는 용도로 사용된 기법이며, IPA는 1977년 Martilla[17]가 자동차 사업의 성취도를 분석하기 위해 다속성 모델을 기초로 하여 중요도와 만족도를 동시에 분석하는 마케팅 기법이다[18]. 해당 연구는 금융 분야의 개인정보 항목을 TF-IDF기법을 통해 측정 후 IPA기법을 적용하여 개인정보의 위험도를 측정하여 다양한 개인정보 항목의 중요도와 성과 분석 모델을 Fig. 3과 같이 제시하였다.

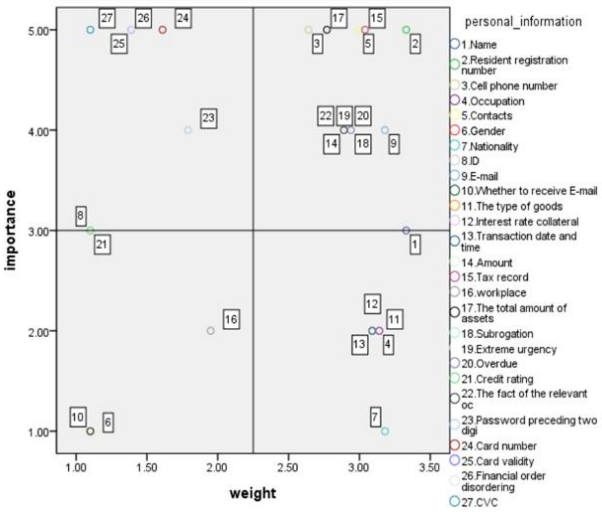

Fig. 3. Importance-Performance Analysis Model

2.3 Personal Information Exposure Risk Using Information Volume

Kim [19]등은 기존에 제시된 개인정보 노출 위험도를 측정하는 방법들이 데이터 상황에 따라 위험도 측정 방법을 달리 수행함으로써 발생하는 오류를 보완하기 위해 정보량을 활용하여 노출 위험도를 정량적으로 측정하는 방안을 제시하고 있다. 제안하는 알고리즘은 1. 각 열의 개인정보 노출 위험도를 측정하고, 2. 여러 열을 고려한 최종 개인정보 노출 위험도를 측정하는 단계로 구성되어 있으며, UCI(University of California, Irvine)의 교육 데이터셋을 통해 검증한 결과 정보량의 개수가 상대적으로 적거나 소수를 차지하는 항목의 경우 노출 위험도 값이 비교적 높게 나타나는 것을 확인할 수 있었다.

2.4 Identification of Personal Information in Interactive Text Data

Kang [20]등은 개인정보 탐지를 위한 특화 개체명 주석 데이터셋을 구축하여 분류하는 실험을 진행하였다. 개인정보 활용 증가에 따른 프라이버시 침해 등 보호의 중요성의 증대에 따라 규칙기반(Rule-based)과 딥러닝(Deep Learning)을 활용하여 일반 개체명 및 정규식 기반의 개인정보 탐지 성능과 특화 개체명을 활용한 탐지를 비교하는 실험으로 후자의 경우 전자의 한계였던 이름, 주소 등 일반 문자열로 구성된 개인정보에 대한 탐지가 가능했으며, 일반 개체명에서 하나의 단위로 잡히지 않았던 비교적 긴 시퀀스 정보에 대해서도 하나의 단위로 탐지될 수 있음을 확인하였다. 다만, 개인정보 특화 태그셋을 7개만 사용한 점과 원천 데이터가 유출 위험 정도(intensity)에 대한 세밀한 주석이 이뤄지지 않은 한계가 보완 사항으로 제시되었다.

3. Implications

지금까지 다양한 선행연구를 통해 개인정보의 재식별 위험성을 낮추기 위해 고려해야 하는 사항과 기술적으로 재식별 위험도를 측정하는 사항에 대해 살펴보았다. 개인정보 재식별 위험을 경감하기 위한 관리적 방안으로는 Elliot [7]이 제시한 데이터 환경(활용되는 분야 및 보호되는 수준)을 고려하여 재식별 가능성을 고려해야 하며, NIST SP800-188[9] 및 ISO/IEC25237[10]에서 제시하고 있는 데이터 공유모델과 개인정보 보증 수준별 가명처리 수준을 고려하여 비식별 정보가 활용되는 기관에서의 개인정보보호 수준이 높은 경우 비식별 조치 수준을 낮춤으로써 데이터의 유용성은 제고하고, 활용 과정에서 재식별 가능성을 낮출 수 있는 방안을 살펴볼 수 있었다. 다만, 대화형 텍스트가 가지고 있는 순차적으로 맥락이 연결되는 특징이나, 다양한 유형의 참여자 의도, 감정 등을 고려한 위험도 측정 방법이 없음을 확인할 수 있었다.

기술적 재식별 측정 방안으로는 PDRA모델[14]을 활용하여 항목별 위험도를 분류할 수 있었으며, 원본 데이터에 등장하는 개인정보의 항목의 빈도를 활용하여 위험도를 측정하는 IPA[17] 방식 및 정보량을 통해 노출 위험도를 측정하는 방법[19]을 살펴보았다. 다만, 기술적 위험성 측정에서는 데이터가 활용되는 환경 및 보호조치에 대한 경감처리, 위험도 측정에 따른 구체적인 처리수준 등은 제시되지 않음을 확인할 수 있었다.

위와 같은 문제점을 해결하기 위해 제3장을 통해 대화형 텍스트 데이터를 기반으로 데이터 환경(Context)을 고려한 새로운 재식별 위험도 측정 방법론을 제시하고자 한다.

III. Proposed Methodology for Measuring Risk of Re-Identification of Personal Information

본 연구에 대한 가설을 세우기 위한 사전 준비로 1. 사전 데이터 분석, 2. 가상의 데이터 구축, 3. 재식별 가능성이 높은 개인정보 항목 도출 단계를 수행하였으며 구체적인 사항은 다음과 같다.

첫번째, 사전 데이터 분석은 기존의 대화형 텍스트 데이터에 대한 정의 및 다양한 주제들에 대한 기준을 수립하는 과정으로 국립국어원의 '메신저 대화 자료 수집 및 말뭉치 구축 사업[21]'의 결과를 활용하여 대화 유형의 분류 및 발화할 수 있는 개인정보 항목을 선정하였다.

두 번째, 가상의 데이터 구축 단계로 대화형 텍스트 데이터의 재식별 위험도를 분석하기 위해서는 개인정보가 포함된 대화형 텍스트 데이터가 필요하다. 하지만 현행법률상 개인정보가 포함된 대화형 데이터는 「개인정보 보호법」에 따라 공개될 수 없으며 별도의 정보 주체의 동의 없이 수집도 불가능한 사항이다. 따라서, 클라우드소싱(Crowdsourcing)을 통해 1단계(사전 데이터 분석)에서 정의한 주제 및 항목을 기준으로 약 8,000건의 가상 대화형 데이터셋을 생성하였으며 이렇게 생성된 가상의 데이터셋을 기준으로 개별 분석을 통해 개인정보 항목별 위험도를 측정하였다.

마지막 세 번째 단계에서는 위에서 도출된 분석 결과를 기반으로 주민등록번호 등 법률에서 사용을 금지하는 항

목을 배제하고 최종 항목을 도출하였다. 구체적인 단계별 세부 사항은 다음과 같다.

1. Preliminary Preparation

1.1 Step 1 (Preliminary Data Analysis)

대화형 텍스트 데이터는 두 명 이상의 참여자 간의 상호작용을 기반으로 생성이 되며 대화는 맥락에 따라 순차적으로 진행되는 특징이 있다. 또한 질의응답, 자유토론, 고객상담 등의 다양한 유형과 참여자의 의도, 감정, 태도 등도 반영될 수 있어 이러한 대화형 데이터의 유형을 정의하기는 매우 어려운 일이다. 이에 본 연구에서는 메신저 대화 특성을 대표성·균형성 있게 반영하고 개인정보 및 저작권 이용 등의 법적 문제가 없는 국립국어원의 '메신저 대화 말뭉치' 자료를 원시데이터로 활용하였다. 해당 보고서에 따르면 메신저 대화의 주제 분류 항목을 16개로 분류하고 주제별 대화 내용에 포함되어 있을 항목을 키워드로 제시하고 있다[21]. 본 연구는 이러한 주제에서 개인정보가 포함된 8가지의 주제를 추출하였으며, 주제별 포함되어 있을 것으로 예상되는 개인정보 항목은 중복을 제외하여 선정하였다. 구체적인 대화 유형의 정의는 Table 1과 같다.

1.2 Step 2 (Building and Analyzing Virtual Data)

1단계에 정의한 Table 1의 개인정보 항목(총 33개)을 기준으로 한국언론진흥재단이 공개한 AI 언어모델인 KPF- BERT[22] 알고리즘을 활용하여 약 54,000개의 개체를 가상으로 구축하였으며, 클라우드소싱 방식으로 이러한 개체명이 태깅된 가상 대화 세트를 약 8,000건 생성하였다. 클라우드소싱은 총 18명의 작업자가 참가하였으며 Table 2와 같이 하나의 대화에 3회~5회의 질문과 응답을 주고받는 Turn이 존재하도록 작성하였다. 또한 한 대화문에 같은 개인정보가 3개 이상 포함되지 않도록 하였으며, 영문 표기는 모두 한글 음차로 변환하고, 실제 일상대화처럼 자연스럽게 이어질 수 있도록 작성하였다.

Table 1. Topic classification items for messenger conversations

Topic of Conversation	Personal Information Items
individual, relationship	Name, Nickname, Birth, Age, Gender, Religion, Nationality, etc. (almost any personal infor mation item may be included)
housing, living	Address, Place name, Vehicle Num
shopping, trading	Name, Birth, Mobile phone/Card/Account number, Email address, ID, URL
public service	RRN, Passport/Driver's/Foreigner, Telephone/FAX number, IP info
leisure, entertainment	Nicknames/Attractions, Clubs, IDs
work, occupation	Job name, Department name, Position
beauty, health	Age, Gender, Height, Weight, Blood type, Medical number
learning, career	Name of school, Grade, Major

Table 2. Example fictitious conversation set (2 times)

A: Sir, are you religious? B: Yes, I'm a Catholic. A: There's always a cross on my desk. B: You have a great eye for detail.

구축한 대화 세트를 분석한 결과 평균적으로 한 대화 세트에 약 6.7개의 개인정보가 태깅되어 있으며, 대화 세트에서 가장 많이 등장하는 개인정보는 이름이었으며, 상대방이나 특정인을 지칭하기 위한 항목(직책/직급, 별명/애칭)과 위치정보(장소명, 주소), 소속에 대한 정보(직장명, 부서명, 동아리/동호회)가 대화에서 자주 등장하였다. 그 외 여권번호, 운전면허번호, IP정보 등은 평균적으로 대화

Table 3. Frequency of personal information utterance in virtually constructed conversation data

No.	Item	Avg	SD	Max	No.	Item	Avg	SD	Max	No.	Item	Avg	SD	Max
1	Name	1.09	1.46	11	12	Address	0.33	0.95	12	23	Vehicle num	0.11	0.42	5
2	Nick names	0.29	0.88	14	13	Place	0.35	0.80	9	24	Job name	0.33	0.74	13
3	Birth	0.14	0.41	5	14	RRN	0.14	0.45	5	25	Depart ment	0.24	0.68	9
4	Age	0.14	0.44	5	15	Foreigner	0.12	0.45	5	26	Position	0.34	0.92	13
5	Gender	0.13	0.45	6	16	Passport	0.10	0.40	4	27	School	0.19	0.60	7
6	Height	0.12	0.44	5	17	Driver's	0.10	0.42	6	28	Grade	0.12	0.43	5
7	Weight	0.12	0.44	5	18	Mobile	0.18	0.45	4	29	Major	0.17	0.62	10
8	Blood	0.11	0.51	9	19	Telephone	0.19	0.47	4	30	ID	0.12	0.44	7
9	Religion	0.15	0.72	10	20	Card	0.17	0.49	5	31	URL	0.12	0.38	3
10	Nation	0.13	0.51	6	21	Account	0.19	0.48	5	32	IP	0.10	0.50	8
11	Clubs	0.30	0.82	10	22	Email addr	0.13	0.41	6	33	Unit	0.12	0.65	11

에서 가장 적게 등장하는 것으로 나타났다. 구축한 데이터의 개인정보 항목에 대한 통계는 Table 3과 같다.

1.3 Step 3 (Deriving Re-identifiable Personal Information Items)

사전준비 단계의 마지막으로 2단계에서 수행한 개인정보 항목을 기반으로 주민등록번호, ID, 카드번호 등 법률 및 국내 가이드라인 등에서 정의[8]하고 있는 식별정보(특정 개인과 직접적으로 연결되는 정보)를 제외하고 다른정보와 결합 또는 연계 시 특정 개인을 식별할 가능성이 높은 항목 19개를 Table 4과 같이 최종적으로 선정하였다.

Table 4. Items that can be re-identified through linkage with other information

No.	Item	No.	Item
1	Name	11	Address
2	Nicknames	12	Place
3	Birth	13	Telephone
4	Age	14	Job name
5	Gender	15	Department
6	Height	16	Position
7	Weight	17	School
8	Blood	18	Grade
9	Religion	19	Major
10	Clubs		

2. Proposed Re-Identification Risk Measurement Methodology

본 연구의 범위는 대화형 텍스트 데이터를 기반으로 데이터에 포함된 개인정보 항목의 재식별 위험도를 계량적으로 산정하는 기준을 마련하는 것이다. 여기서 재식별이란 특정 개인을 알아볼 수 없도록 비식별 처리된 정보가 다른 정보 등과의 결합·연계를 통해 특정 개인을 식별할 수 있는 것을 의미하고 있다. 이러한 재식별의 위험성을 판단하기 위해 Kim [23]은 비식별 데이터 자체만의 위험성뿐만 아니라 데이터가 이용되는 환경, 즉 데이터 환경(Context)을 검토해야 한다고 제시하고 있으며, 국내외의 비식별 조치 관련 가이드 등에서도 데이터의 활용 또는 이용 환경을 종합적으로 검토해야 한다고 제시하고 있다.

이에 본 연구에서 제안하는 방법론은 기본적으로 데이터 활용환경(A)과 데이터 자체의 위험도(Si)를 합산하고, 최종 데이터가 이용되는 환경에 따라 위험도(B)를 경감하는 방식으로 계산식은 Fig. 4와 같다.

$$S = A + \sum_{i=1}^{n} Si - B$$

Fig. 4. Proposed re-identification risk calculation formula

데이터 활용 환경(A)의 경우 3개로 구분하여 가명정보를 제공받는 제3자가 특정되어 있는 경우와 불특정 되어 있는 경우, 그리고 데이터가 완전히 공개되는 환경을 고려하였다. 각 구분별 위험도를 최고 5점을 부여하여 현행 법률의 테두리 안에서 활용되고 있는 가명정보의 경우 1점, 제공받는 제3자가 특정이 되지 않지만, 정부에서 인정된 '개인정보 안심구역' 등 안전한 장소에서 가명정보를 활용하는 경우 3점, 공개의 경우 데이터를 활용하는 자 및 데이터의 안전한 관리가 이뤄질 수 없는 점[24]을 고려하여 5점을 부여하였다. 위험도 점수가 높게 부여되는 경우 재식별 가능성이 높은 환경으로 정의하였다.

데이터 자체의 위험도의 경우 ①데이터 자체의 통계적 위험성(Statistics)과 ②시간의 경과에 따라 값이 변경될

Table 5. Risk calculation standards by interactive data-based characteristics

Statistics			Rigidity			Recency		
No.	Item name	Score	No.	Item name	Score	No.	Item name	Score
1	Name, Age, Birth	0.9	1	Birth, Blood type	1.0	1	Name, Birth, Gender, Blood type, Tel/Fax Num	1.0
2	Nick names	0.8	2	Name, Gender	0.9	2	Religion	0.9
3	Gender, Position, Place name	0.7	3	Tel/Fax Num, Religion, Major	0.8	3	Address, Place name	0.8
4	Job/School name	0.6	4	Address, Place/School name	0.7	4	Nickname	0.7
5	Department name, Grade, Major	0.5	5	Job/Department name, Grade	0.6	5	Job/School name	0.6
6	Tel/Fax Num, Address	0.4	6	Nickname, Club, Position	0.5	6	Department name, Grade	0.5
7	Religion, Club	0.3	7	Age	0.4	7	Position, Major	0.4
8	Height, Weight	0.2	8	Height	0.3	8	Age, Club	0.3
9	Blood type	0.1	9	Weight	0.2	9	Height	0.2
						10	Weight	0.1

가능성이 작아 개인을 재식별할 위험성이 높은 경직성(Rigidity), ③시간의 흐름에 따라 데이터가 자주 변경되는 특성이 있는 최신성(Recency), ④전체 데이터의 고유값 또는 편중된 분포를 통해 재식별 위험성이 높은 특이정보 등 4가지의 위험도를 검토하여 각각 위험도 점수를 산정할 수 있도록 하였다. 통계성, 경직성, 최신성의 경우 다양한 데이터 분석 경험이 있는 전문가 3인과 함께 4차례의 검토 회의를 거쳐 점수를 부여하였으며, 특이정보의 경우 통상적으로 인지하고 있는 항목은 1점, 특이정보가 있지만 전처리를 하는 경우 또는 특이정보가 거의 없거나 노출이 되어도 특정 개인에 대한 프라이버시 침해 영향이 거의 없는 경우 0.1점을 산정하였다. 최종적으로 산출된 특성별 데이터 위험도 산정 기준은 Table 5와 같다.

마지막으로 데이터를 이용하는 처리자의 보호 환경이 강할수록 정보의 유출이나 외부 공격 등을 통한 재식별 가능성을 낮출 수 있다. 따라서, 데이터 이용환경이 개인정보보호 법률에서 정한 수준보다 높게 보호하는 경우(개인정보보호 인증서 취득 등) 0.5점을 부여하고, 법률의 규정 사항만을 준수하는 경우 0.3점, 법률 규정 사항보다 낮은 수준으로 보호하는 경우 0.1점을 부여하고 정보를 공개하는 경우 보호 수준이 없는 것으로 0점을 부여하였다.

위와 같이 제안하는 방법론을 활용한 개인정보 항목별 위험도 측정에 대한 예시는 Fig. 5와 같다. 해당 데이터가 향후 A사와의 계약을 통해 제공되며 하나의 대화형 데이터에는 이름, 주소, 나이, 성별의 4가지 '개인식별 가능정보'가 포함되어 있으며, 특이치에 대해서는 전처리를 통해 사전에 제거를 하였다. A사는 ISMS-P등 개인정보를 안전하게 활용할 수 있는 환경과 인증을 취득하였으며, 가명정보 처리에 관한 책임자도 별도로 지정하여 운영 중이다. 이러한 경우 데이터 활용 환경은 1점을 부여하고, 데이터 위험도의 경우 하나의 대화 내의 '개인식별 가능정보'의 포함 개수만큼 통계성, 경직성, 최신성의 위험을 합산한다. 마지막으로 데이터 이용환경이 안전하기 때문에 0.5점을 경감하여 최종 9.8점을 부여할 수 있다.

$$Risk = 1 + ((0.9 + 0.9 + 1.0 + 0.1) + (0.4 + 0.7 + 0.8 + 0.1) + (0.9 + 0.4 + 0.3 + 0.1) + (0.7 + 0.9 + 1.0 + 0.1)) - 0.5 = 9.8$$

Fig. 5. Example of proposed risk measurement

해당 연구에서는 위험도에 대한 임계값 기준을 8로 설정하고 있는데 이러한 이유는 3.1. 사전준비 단계에서 구축한 가상 대화 데이터의 분석 결과를 반영하였으며 데이터 위험도만을 기준으로 하나의 대화 세트에 6개의 개인정보 항목이 포함되는 경우 나올 수 있는 위험도 측정값은 0~17.4로 50%를 기준으로 재식별 위험도를 판단하였다. 이러한 임계치 기준은 실무에서 위험을 어느 정도로 수용할지에 따라 유동적으로 설정할 수 있다. Fig. 5를 통해 예시로 제시하였던 대화 데이터의 경우 임계치 범위를 초과하기 때문에 개인식별 가능성에 대한 추가 검토를 권고하여 향후 발생할 수 있는 재식별 가능성을 사전에 검토할 수 있다.

IV. Validation of Proposed Methodology

제안하는 방법론의 데이터 위험도 산정 기준 및 측정 방법에 대한 타당성 검토를 위해 첫 번째로 제3장에서 구축한 가상의 데이터를 이용하여 자체적인 실증을 수행하였으며, 두 번째로 제안 방법론에 대한 구체적인 설명과 측정 예시를 포함하여 외부 전문가 총 220명을 대상으로 설문조사를 실시하였다.

1. Self-Validation

자체 실증의 경우 제3장에서 구축한 약 8,000건의 개인정보가 포함된 가상의 대화형 텍스트 데이터를 이용하여 무작위로 400개의 데이터를 추출하였으며, 추출한 데이터를 기반으로 제안하는 방법론을 적용하여 위험도를 측정하였다. 실증 결과 이름, 별명/애칭, 생년월일, 나이, 성별, 키, 몸무게 등의 정보가 포함되었을 때 위험도가 높은 수준으로 나왔으며, 전공, 학년, 학교명, 직책/직급, 부서명, 장소명 등의 정보가 포함된 경우는 상대적으로 낮은 위험도가 측정되었다. 또한, 제안한 측정방법론의 편향 위험 감소 및 데이터 자체 위험도 점수의 적절성을 검토하기 위해 400개의 데이터를 위험도가 높은 항목이 포함된 데이터와 위험도가 낮은 항목이 포함된 데이터를 구분하여 비교한 결과, 위험도가 높은 항목의 경우 4개 이상의 개인정보 항목이 포함되었을 때 임계치를 초과하였으며, 위험도가 낮은 항목의 데이터의 경우 5개 이상의 개인정보 항목이 포함된 경우만 임계치를 초과한 것을 Table 6과 같이 확인할 수 있었다.

Table 6. Risk analysis for each item based on virtual data

		high-risk data	Low-risk data
MIN		7	3.1
MAX		14.3	9.1
Mean		10.1	5.8
Distribution (PI Items in Data)	3	-	148
	4	122	50
	5	75	2
	6	3	-

2. Survey Result

외부 전문가를 대상으로 실시한 설문조사의 경우 총 220명의 기업·기관의 개인정보보호 담당자 및 개인정보보호 분야 전문가, 개인정보 관련 데이터처리 전문가 등이 응답해 주었다. 구체적인 설문조사의 결과는 Table 7, Table 8, Table 9과 같다.

Table 7. Validity of Virtual Data Construction and Analysis Results

No	Question	M	SD
1-1	Is it appropriate to use AI to construct virtual conversational text data containing personal information for measuring 're-identification risk'?	4.77	0.42
1-2	Are the analysis results, which derive re-identifiable personal information items from the virtual conversational data created in 1-1, valid?	4.64	0.48

Table 8. Validity of the Proposed Methodology

No	Question	M	SD
2-1	This study measures the risk of personal information items not only in conversational text data itself but also by examining the data's use and environment comprehensively. Do you consider the proposed risk measurement method valid?	4.64	0.49
2-2	Do you believe that the method and scoring of assigning risk levels based on the usage environment of conversational text data are appropriate?	4.49	0.58
2-3	In the case of data risk, we review the risk of each personal information item based on ① statistical risk of the data itself, ② rigidity with no change in value over time, and ③ recency with frequent changes of data over time. Are the risk score calculation criteria for each item appropriate?	3.5	1.03
2-4	The data usage environment (privacy level) is reviewed to provide a procedure to mitigate risk. Do you think these standards are appropriate?	4.37	0.64
2-5	A practical example using the proposed risk measurement methodology is presented in [Overview]. Do you think this overall risk metric is appropriate?	3.81	0.66

Table 9. Effectiveness of Research

No	Question	M	SD
3-1	Recently, there has been a growing demand for the use of interactive text. What do you think about the standards for determining whether such interactive text data contains personal information and assessing risk?	4.5	0.5
3-2	If the original technology is developed using the proposed methodology, do you think it can help create an ecosystem for safe use of interactive text data containing personal information in the future?	4.39	0.49
3-3	If this technology becomes widespread in the future, would you be willing to introduce or use it in your company/organization?	4.27	0.44

Table 7인 제안하는 방법론 및 위험도 측정을 위해 AI를 활용한 가상의 대화형 데이터를 구축한 절차 및 분석 결과의 경우 Cronbach α값 0.843, 표준편차의 평균 0.45로 높은 수준의 신뢰도를 나타냈으며, AI로 가상의 대화형 텍스트 데이터를 구축하여 주요 개인정보 항목을 도출하고, 가상의 데이터 분석 결과를 통해 재식별 가능한 개인정보 항목을 도출하는 절차 및 결과가 매우 타당하다는 결과를 도출할 수 있었다. Table 8의 제안하는 위험도 측정 방법론에 대한 타당성 조사 결과는 Cronbach α값 0.779, 표준편차의 평균은 0.68로 기본적인 수준의 신뢰도를 나타냈으며, 비교적 낮은 점수가 도출된 2-3 항목의 경우 각 항목별로 부여하는 위험도 점수 기준에 대한 근거가 명확하지 않다는 의견이 있었다. 이 부분에 대해서는 향후 해당 연구 기반의 솔루션을 개발하여 현업을 대상으로 실증을 통해 다양한 사례를 도출하여 기준 점수에 대한 레퍼런스를 제공할 예정이며, 2-5 항목의 경우 대화형 텍스트 데이터에 포함된 개인정보의 항목이 많은 경우 위험도는 높아질 수 밖에 없다는 의견이 있었다. 본 연구의 범위는 3.1.2에서 구축한 데이터셋이 3~5회 정도의 대화를 기준으로 산정한 방법론이기 때문에 향후 대량의 개인정보 항목이 포함되어 있는 경우 제안하는 알고리즘을 개선할 필요성이 나타났다. 다만, 대화형 데이터 특성 상 맥락에 대한 구분과 이해 없

이 무작정 대량의 데이터를 학습시키는 경우 결과물로 도출된 AI챗봇의 성능의 정확도는 떨어질 수 밖에 없을 것이기에 매우 정교한 전처리 작업이 필요할 것으로 판단된다. 그 외의 위험도 측정방법에 대해서는 전반적으로 타당성이 높다는 결과는 도출할 수 있었다. 마지막으로 Table 9의 연구 효과성에 대한 타당성 조사 결과는 Cronbach α값 0.777, 표준편차의 평균은 0.48로 대화형 텍스트 데이터의 활용 수요 증가와 함께 제안하는 방법론의 필요성이나 데이터 활용 환경 개선, 향후 솔루션 도입 의사 등 모든 항목에서 높은 효과성이 있음을 알 수 있었다.

V. Conclusions

2020년 데이터 3법 시행에 따라 개인정보를 안전하게 가명처리하여 AI학습 등의 빅데이터로 활용할 수 있는 근거가 마련되었으나, 학습된 데이터의 전처리 미흡 및 개인정보 재식별 가능성을 고려하지 않은 데이터 처리 등으로 인해 AI챗봇 등 빅데이터를 이용한 신산업에서 프라이버시 침해 사례가 자주 등장하고 있다. 정부는 이러한 침해 사례가 빅데이터를 활용하기 위한 법, 제도의 실효성 문제나 관련 산업의 위축 등을 우려하고 있으며, 이를 보완하기 위한 다양한 개인정보 강화 기술(PET : Privacy Enhancing Tech nology) R&D를 추진하고 있다.

이에 본 연구는 정부 정책의 일환으로 개인정보가 포함된 비정형 대화형 텍스트 데이터를 기반으로 '데이터 자체의 특성'과 '데이터 환경(Context)'을 고려하여 향후 발생할 수 있는 개인정보의 재식별 위험도을 계량적으로 측정할 수 있는 방법론을 제안하였으며, 이러한 방법론을 활용하여 사전 학습에 이용되는 개인정보(식별자 삭제 및 준식별자 포함)를 대상으로 위험도를 측정하고, 데이터를 활용하는 기업이 요구하는 위험도 임계치에 따라 측정된 항목의 가명처리 수준을 재검토(추가 가명 처리)할 수 있는 방법론이 될 수 있을 것이다. 본 연구는 비정형 데이터의 향후 활용하는 과정에서 발생할 수 있는 재식별 가능성을 사전에 검증할 수 있는 유일한 방법론이란 점에 의의가 있다.

마지막으로 AI모델에 양질의 학습데이터가 없다는 것은 마치 재료없이 요리하려는 요리사와도 같다. AI를 제대로 학습시키기 위해서는 많은 양의 데이터를 확보하는 것도 중요하지만, 그것보다 초기부터 '좋은 데이터'를 확보하는 것이 중요하며, 이러한 데이터에 특정 개인이 식별되어 프라이버시 침해가 발생하지 않도록 조치를 하는 것도 필수적인 요소이다. 후속 연구로는 제안 방법론의 타당성 검증을 통해 제시되었던 의견을 반영하여 현재 3~5회 정도의 대화를 기준으로 측정하는 방법론을 확장하여 대량의 대화 데이터셋을 기반으로 본 방법론을 적용하여 위험도를 측정할 수 있는 방법에 대한 연구를 수행할 예정이며, 추가로 본 방법론을 적용한 다양한 대화유형의 실증을 통해 국내 정부의 지침 등에 반영하여 신기술을 보급할 수 있도록 노력할 예정이다.

ACKNOWLEDGEMENT

This study was funded by the Personal Information Protection Commission of the Republic of Korea and the Korea Internet & Security Agency (KISA), grant number 1781000017.

REFERENCES

[1] Jun-ho Park, Artificial Intelligence-Based Chatbot System Technology Trend, Korea Information Processing Society, Vol.26, No.2, pp39-46, 2019. UCI : I410-ECN-0102-2019-500-001455843

[2] Avinash Chandra Das et al, The next frontier of customer engagement: AI-enabled customer service, Mckinsey& Company, 2023.3. https://mck.co/40y0s9A/

[3] Chatbot Market Global Industry Analysis, Precedence Research, 2023. https://www.precedenceresearch.com/chatbot-market

[4] Xiaodong Wu et al, Unveiling Security, Privacy, and Ethical Concerns of ChatGPT, Journal of Information and Intelligence, Vol.2, Issue 2, pp.102-115, 2024. https://doi.org/10.1016/j.jiixd.2023.10.007

[5] Seung-Jae Jeon, Possibility of Using Personal Information as Machine Learning Data Seen Through the Iruda Case, Korea Association For Info-Media Law, Vol.25, No.2, pp.103-133, 2021. https://doi.org/10.22846/kafil.25.2.202108.004

[6] Heui-ok Lee, Ethical Guidelines for Controlling Bias of Artificial Intelligence Chatbots, Korean Public Law Association, Vol.51, No.3, 2023.2. http:/doi.org/10.38176/PublicLaw.2023.2.51.3.715

[7] Mark Elliot, Elaine Mackey et al, "The Anonymisation Decision-making Framework", UK Anonymisation Network, May. 2016.

[8] Personal Information Protection Commission, "Guidelines for processing Pseudonymization information", Sep. 2020.

[9] Simson Garfinkel, NIST800-188 De-Identifying Government Data Sets, NIST, 2022. https://doi.org/10.6028/NIST.SP.800-188

[10] ISO/IEC 25237, Health informatics- Pseudonymization, 2017.

https://www.iso.org/standard/63553.html
[11] Personal Information Protection Commission, “Guidelines for Personal Information Impact Assessment”, Apr. 2024.
[12] Personal Information Protection Commission, “Personal information risk analysis standards and commentary”, 2020.
[13] Ministry of the Interior and Safety, guidelines for homapage personal Information exposure prevention, 2017.
[14] Eu-gene Kim, Privacy Detection and Risk Analysis Model, Sungshin Women's University, Master's Thesis, 2011.
[15] Su-jun Jeong, A Study on Analysis of Personal Information Risk Using Importance-Performance Analysis, The Journal of The Institute of Internet, Vol.15, No.6, pp.267-273, 2015. http://dx.doi.org/10.7236/JIIBC.2015.15.6.267
[16] Sung-jick Lee, "Keyword Extraction from News Corpus using Modified TF-IDF", Journal of Society for e-Business Studies, 14(4):59-73, 2009.
[17] J. A. Martilla and J. C. James, “Importance Performance Analysis”, Journal of Making, Vol.41, pp. 77-79, 1977.
[18] Jo-seong lae, “ A Study of the Aged in the Leisure Life of Leisure Motivation and on the Leisure Satisfaction”, Honam for master's thesis. 2013. 2
[19] Chae-hyeon Kim, An Information Content-based Method for Measuring the Risk of Personal Information Exposure, Korean Institute of Information Scientists and Engineers, pp. 926-928, 2022.
[20] Hye-rin Kang, A Study on the Construction of Specialized NER Dataset for Personal Information Detection, Annual Conference on Human and Language Technology, pp. 185-191, 2022.
[21] National Institute of Korean Language, “Messenger Corpus”, 2022.
[22] KOREA PRESS FOUNDATION, Understanding BERT in the history of artificial intelligence, 2022.
[23] Sweeney L, Re-identification Risks in HIPAA Safe Harbor Data, PubMed Cenral, (2017). https://techscience.org /a/2017082801
[24] Dong-hyun Kim, A study on Data Context-Based Risk Measurement Method for Pseudonymized Information Processing, Journal of The Korea Society of Computer and Information, 2022, Vol.27, No.6, pp.53-63. https://doi.org/10.9708/jksci.2022.27.06.053

Authors

Dong-Hyun Kim received Ph.D degree in convergence security from Chung-Ang University, Korea, in 2022. He has been conducting personal information surveys and policy improvement for 6 years from 2010, and since 2016, the Data Utilization Support Team has been working to utilize safe personal information as big data. He is interested in Personal Information Security, De-Identification & De-Identified Information Risk Measure.

Ye-Seul Cho received the B.S. degree in Statistics from Sungshin Women’s University, Korea, in 2018. She has been working as a senior researcher at World Vertex. She is interested in Natural Language Processing, Personal Information Protection.

Tae-Jong Kim received his master's degree in e-learning from the Korea National Open University Graduate School in 2020. He has been conducting research in the field of edutech artificial intelligence since 2017, and has been conducting research to utilize safe personal information as big data since 2022. He is the CEO of Worldvertex, an edutech solution provider, and is conducting various research related to personal information protection.