applied sciences

MDPI

Article

Automated Redaction of Personally Identifiable Information on Drug Labels Using Optical Character Recognition and Large Language Models for Compliance with Thailand's Personal Data Protection Act

Parinya Thetbanthad, Benjaporn Sathanarugsawait and Prasong Praneetpolgrang *

School of Information Technology, Sripatum University, Bangkok 10900, Thailand; parinya.the@spumail.net (P.T.); benjaporn.sa@spu.ac.th (B.S.)
* Correspondence: prasong.pr@spu.ac.th

Abstract: The rapid proliferation of artificial intelligence (AI) across various industries presents both opportunities and challenges, particularly concerning personal data privacy. With the enforcement of regulations like Thailand's Personal Data Protection Act (PDPA), organizations face increasing pressure to protect sensitive information found in diverse data sources, including product and shipping labels. These labels, often processed by AI systems for logistics and inventory management, frequently contain Personally Identifiable Information (PII). This paper introduces a novel AI-driven system for automated PII redaction on label images, specifically designed to facilitate PDPA compliance. Our system employs a two-stage pipeline: (1) text extraction using a combination of EasyOCR and Tesseract OCR engines, maximizing recall for both Thai and English text; and (2) intelligent redaction using a pre-trained large language model (LLM), Qwen (Qwen/Qwen2.5-72B-Instruct-AWQ), prompted to identify and classify text segments as PII or non-PII based on simplified PDPA guidelines. Identified PII is then automatically redacted via black masking. We evaluated our system on a dataset of 100 drug label images, achieving a redaction precision of 92.5%, a recall of 83.2%, and an F1-score of 87.6%, with an over-redaction rate of 3.1%. These results demonstrate the system's effectiveness in accurately redacting PII while preserving the utility of non-sensitive label information. This research contributes a practical, scalable solution for automated PDPA compliance in AI-driven label processing, mitigating privacy risks and promoting responsible AI adoption.

Keywords: PDPA; data privacy; data protection; PII redaction; label redaction; image redaction

Received: 18 February 2025
Revised: 11 April 2025
Accepted: 17 April 2025
Published: 29 April 2025

Citation: Thetbanthad, P.; Sathanarugsawait, B.; Praneetpolgrang, P. Automated Redaction of Personally Identifiable Information on Drug Labels Using Optical Character Recognition and Large Language Models for Compliance with Thailand's Personal Data Protection Act. *Appl. Sci.* **2025**, *15*, 4923. https://doi.org/10.3390/app15094923

1. Introduction

The transformative influence of artificial intelligence (AI) is rapidly reshaping industries and societies globally. It drives innovation and efficiency across diverse sectors, including healthcare, finance, manufacturing, logistics, and customer service [1]. As AI systems become more sophisticated and integrated into data-driven processes, the volume and velocity of processed data have reached unprecedented levels.

This data-centric paradigm fuels significant advancements. However, it also concurrently amplifies legitimate concerns surrounding personal data privacy and security [2]. These concerns are particularly relevant in contexts where seemingly innocuous data sources, like labels on products and packages, can contain sensitive information. In

this era of ubiquitous data collection and processing, safeguarding personal information has evolved from an ethical consideration to a core principle of responsible technological development and deployment. This is mandated by increasingly stringent legal frameworks worldwide [3].

Reflecting this global trend, Thailand enacted the Personal Data Protection Act (PDPA) in 2019, with full enforcement beginning in 2022. This landmark legislation aims to protect the personal data of individuals within Thailand, aligning with international best practices for data privacy [2,4]. The PDPA mandates that all organizations operating in Thailand, regardless of sector, size, or business nature, must implement effective measures to ensure the privacy, security, and lawful processing of personal data. This obligation extends to all data processing activities, encompassing the collection, storage, use, and dissemination of personal information.

Crucially, this mandate includes data derived from various sources, explicitly including personal data embedded within or extracted from sources such as labels. These labels often contain personally identifiable information (PII) like names, addresses, and contact details. Compliance with the PDPA is therefore not merely a legal obligation; it has become a critical factor for organizations operating in Thailand. It is linked to maintaining public trust, fostering ethical data stewardship, ensuring the sustainability of AI-driven business models, and safeguarding brand reputation. Failure to adhere to the PDPA's stipulations and principles can expose organizations to significant financial penalties, legal repercussions, and irreparable damage to their brand image and customer trust [5]. This highlights the critical need for robust, automated solutions to ensure PDPA compliance, especially in data-intensive areas like AI-driven label processing [6].

To address this need, this paper introduces a novel AI-driven system for automated personal data redaction on labels, specifically designed for PDPA compliance. Manual redaction, while a traditional approach, is labor-intensive, time-consuming, and error-prone, making it unsuitable for high-volume label processing [7]. Our system, in contrast, leverages the combined power of Optical Character Recognition (OCR) and large language models (LLMs) to achieve efficient, accurate, and scalable data redaction [8–10].

The key contributions of this work are as follows:

1. An Integrated OCR and LLM Approach: We present a system that combines OCR [9,10] and LLMs [8] for effective data extraction and redaction. This combined approach leverages the strengths of both technologies: OCR for text extraction from images and LLMs for nuanced understanding and identification of PDPA-violating information.
2. PDPA-Specific Violation Detection: We utilize a pre-trained LLM (Qwen/Qwen2.5-72B-Instruct-AWQ) [8] specifically prompted to identify potential violations of the Thai PDPA within the extracted text. This goes beyond simple keyword matching, enabling the detection of context-dependent PII [11].
3. Automated Redaction: The system automatically redacts identified violations directly on the original image, providing a practical solution for data anonymization. This eliminates the need for manual, error-prone redaction processes [7].
4. Empirical Evaluation: We provide a detailed evaluation methodology and present experimental results to demonstrate the system's effectiveness in identifying and redacting PDPA-violating information.

The remainder of this paper is structured as follows: Section 2 reviews related work, Section 3 details the methodology of our AI-driven redaction system, Section 4 presents the evaluation framework and experimental results, Section 6 discusses the findings and limitations, and Section 7 concludes the paper and outlines future research directions.

2. Related Work

To contextualize the contributions of this research and position our AI-driven label redaction system within the broader academic and technological landscape, this section provides a comprehensive review of related work. This review is structured around three key areas directly relevant to our proposed system: Optical Character Recognition (OCR) for labels, privacy-preserving techniques applicable to image and text processing, and the burgeoning field of AI and legal compliance, with a particular focus on data privacy regulations such as PDPA. By examining existing research initiatives, established methodologies, and state-of-the-art technologies in these domains, we aim to highlight the novelty and significance of our proposed approach. We also clearly delineate the specific areas where our work builds upon, extends, and potentially surpasses prior research in the pursuit of automated label redaction for enhanced personal data protection and PDPA compliance.

2.1. Optical Character Recognition (OCR) for Labels

Optical Character Recognition (OCR) technology, a cornerstone of automated text extraction from images, has undergone remarkable advancements in recent decades. These advancements have transformed OCR from a niche technology into a widely adopted tool for diverse applications, ranging from document digitization and archival to automated data entry and information retrieval. A plethora of OCR engines are currently available, each possessing varying strengths and weaknesses in terms of accuracy, speed, language support, and robustness to image variations. These engines span the spectrum from sophisticated commercial solutions, often incorporating proprietary algorithms and extensive training datasets, to accessible open-source tools, fostering wider accessibility and community-driven development.

Among the prominent open-source OCR engines, Tesseract OCR [10] and EasyOCR [9] stand out as particularly influential and widely utilized. EasyOCR, in particular, has garnered significant attention and adoption due to its user-friendly interface, extensive multilingual capabilities, and demonstrated robustness across a range of image conditions [9]. Its support for a substantial number of languages, including complex scripts like Thai, makes it especially relevant to our research, which focuses on PDPA compliance within the Thai context.

The application of OCR technology has been extensively explored across numerous domains, each presenting unique challenges and requirements. Document analysis [12], for instance, has long been a driving force in OCR research, with applications in automated document processing, archival systems, and digital libraries. License plate recognition [13] represents another prominent application area, demanding high accuracy and speed under varying environmental conditions for traffic management, security, and law enforcement purposes. More recently, the application of OCR to product label reading [14] has gained increasing attention, driven by the growth of e-commerce, automated retail, and supply chain optimization.

Extracting information from product labels, however, presents a distinct set of challenges. Labels often exhibit significant variations in font styles, sizes, and layouts, reflecting diverse branding strategies and regulatory requirements. Furthermore, label images can be subject to noise, distortion, and low resolution, particularly when captured in real-world scenarios using mobile devices or automated scanning systems. For languages with complex scripts, such as Thai [15], the challenges are further amplified due to the intricate nature of the script, diacritical marks, and potential ambiguities in character segmentation and recognition. Our research directly addresses these challenges by leveraging EasyOCR, known for its Thai language

support, and by focusing on the specific context of label text recognition for PDPA compliance, where accuracy and reliability are of paramount importance.

2.2. Privacy-Preserving Techniques in Image and Text Processing

The escalating concerns surrounding data privacy in the digital age have spurred significant research and development in privacy-preserving techniques across various data modalities, including images and text. As AI systems increasingly process and analyze visual and textual information, the need to anonymize or redact sensitive data has become a critical imperative. This need is driven by both ethical considerations and increasingly stringent regulatory mandates. A diverse array of techniques has emerged to address this need, each offering different trade-offs between privacy protection, data utility, and computational overhead.

For image data, common privacy-preserving methods typically involve obscuring or removing regions of interest that are likely to contain sensitive information. Blurring techniques, for example, apply Gaussian filters or similar smoothing operations to redact faces, license plates, or other identifiable features [16], effectively reducing the visual discernibility of these regions. Masking approaches, conversely, involve overlaying solid blocks or opaque shapes to completely conceal sensitive areas within an image. Pixelization, another technique, reduces the resolution of specific image regions, rendering details unrecognizable while preserving the overall context.

For text, distinct anonymization techniques exist. Tokenization can replace sensitive words (names, addresses) with generic placeholders. Pseudonymization uses artificial identifiers, allowing analysis while obscuring direct links to individuals. Redaction, the most direct method, involves removing or replacing sensitive text segments, often with black bars or placeholders [11]. Our work primarily employs this redaction (masking) approach. We selected masking for this initial study because of its direct applicability to removing specific PII strings identified by the LLM within unstructured image data, aligning with a strict interpretation of removing defined PII categories under PDPA.

However, the field of privacy preservation is rapidly evolving, offering alternative paradigms. K-anonymization, for example, is a prominent technique primarily used for structured data. Its core principle is to prevent re-identification by ensuring any individual's record is indistinguishable from at least k-1 other records within the dataset. This is typically achieved through generalization (e.g., replacing specific ages with broader age ranges, or full addresses with just the city) or suppression (removing certain identifying values altogether) [17]. Recent advancements focus on optimizing these processes, especially in the context of machine learning challenges, using techniques like iterative local search [18]. Crucially, k-anonymization contrasts with our masking approach: instead of completely removing information, it aims to strike a balance between data privacy and utility by retaining partial, generalized information. This reduces identification risk while potentially preserving more analytical value. While applying k-anonymization directly to unstructured label images presents unique challenges, it represents an important alternative privacy-utility trade-off focused on indistinguishability. Other advanced techniques like differential privacy add calibrated noise to data or query results to protect individuals while enabling aggregate analysis [19].

In the specific context of labels, blanket anonymization that obscures entire labels renders them useless. Therefore, intelligent and selective techniques are essential. Our approach focuses on targeted redaction of identified PII to preserve non-sensitive information utility. While AI techniques like differential privacy and federated learning are increasingly explored for protecting data during model training [20], our research leverages AI (specifically LLMs) for the distinct task of applying redaction to the data itself. Acknowledging

the value of alternative methods like k-anonymization for different scenarios, exploring the adaptation of such concepts to the specific challenges of label image redaction remains an interesting area for future investigation, as noted in Section 7.

2.3. Legal Compliance and AI

The intersection of artificial intelligence and legal compliance has rapidly emerged as a critical and dynamic field. This emergence is propelled by the proliferation of data privacy regulations worldwide. The advent of comprehensive legal frameworks such as the General Data Protection Regulation (GDPR) in Europe [2], the California Consumer Privacy Act (CCPA) in California [21], and, most pertinently to our research, the Personal Data Protection Act (PDPA) in Thailand [4], has fundamentally reshaped the landscape of data processing and privacy protection. These regulations, while varying in specific details, share a common objective: to empower individuals with greater control over their personal data and to mandate organizations to implement robust safeguards. These safeguards protect data from unauthorized access, misuse, and breaches. Compliance with these regulations is no longer merely a matter of adhering to legal mandates; it has become an essential component of responsible data governance, ethical AI practices, and maintaining the trust of stakeholders, including customers, partners, and regulatory bodies.

AI technologies themselves are increasingly recognized as holding significant potential for automating and streamlining various aspects of legal compliance. AI-powered tools can be deployed to assist in data discovery, enabling organizations to efficiently identify and catalog the vast repositories of data they hold. This is a crucial step in understanding data privacy obligations. Data classification, another area where AI excels, can be leveraged to automatically categorize data based on sensitivity levels and regulatory requirements, facilitating targeted privacy controls. Access control mechanisms, critical for ensuring data security and limiting unauthorized access, can be enhanced through AI-driven systems that dynamically manage permissions and monitor data access patterns.

In the specific context of PDPA compliance, AI can play a pivotal role in automating tasks such as identifying personal data within datasets, managing user consent for data processing, and continuously monitoring data security posture to detect and mitigate potential vulnerabilities [22]. However, the application of AI to legal compliance is not without its inherent challenges. Legal regulations, by their nature, are often expressed in complex, nuanced, and sometimes ambiguous language, requiring sophisticated interpretation and contextual understanding. Furthermore, the dynamic and evolving nature of legal landscapes necessitates continuous adaptation and refinement of AI-driven compliance systems. Large language models (LLMs), with their remarkable capabilities in natural language understanding and generation, offer a promising avenue for addressing these challenges. LLMs possess the potential to interpret complex legal documents, extract relevant information, and assist in automating compliance-related tasks that require sophisticated language processing and reasoning [23]. Our research directly explores this potential by leveraging LLMs to interpret PDPA guidelines and make informed decisions regarding the redaction of personal data on labels. We aim to bridge the gap between AI capabilities and the practical requirements of legal compliance in the realm of data privacy.

3. Methodology

Our proposed AI-driven label redaction system is architected as a two-stage pipeline, designed for efficient and accurate personal data redaction from label images. The system strategically combines EasyOCR and Tesseract for text extraction and an LLM for intelligent redaction decisions, ensuring PDPA compliance.

3.1. Stage 1: Text Extraction with EasyOCR and Tesseract

The first stage leverages both EasyOCR [9] and Tesseract OCR [10]. These are open-source OCR libraries known for their multilingual support, including Thai, and robustness. The system processes input label images using both EasyOCR and Tesseract to detect and recognize text. The outputs from both engines are combined, providing both the extracted text and the bounding box coordinates for each text segment. This dual-engine approach is crucial for maximizing text capture accuracy from labels. It helps mitigate potential errors or omissions from a single engine, especially given variations in font, layout, or image quality. We configured both EasyOCR and Tesseract with the Thai language model for Thai labels. Default settings were used for text detection and recognition to ensure broad applicability and ease of use. While both engines are robust, we acknowledge potential challenges with low-resolution images or complex label designs, which we address through pre-processing and error analysis in our evaluation. Combining the results allows us to leverage the strengths of each engine and improve overall recall.

3.2. Stage 2: Redaction Decision with LLM

The second stage employs a large language model (LLM) to analyze the combined text extracted by EasyOCR and Tesseract. The LLM makes intelligent redaction decisions based on PDPA guidelines. We utilize Qwen (Qwen/Qwen2.5-72B-Instruct-AWQ) [8] via its API, accessed using the 'litellm' library [24], leveraging its strong natural language understanding and instruction-following capabilities. The LLM is prompted to classify each extracted text segment as either "personal data" or "not personal data" according to simplified PDPA guidelines.

It is important to note that in this prototype setup, these text segments are transmitted to the external LLM API. While standard security protocols (e.g., HTTPS) are typically used for such transmissions, deploying this system in a production environment handling sensitive data under PDPA would necessitate a thorough security review of the API provider's practices and potentially alternative deployment models, as discussed further in Section 7. If a text segment is classified as "personal data", the system applies black masking focusing on categories like names, addresses, and phone numbers. The prompt engineering process involved iterative refinement to optimize the LLM's accuracy and reliability in identifying personal data. If a text segment is classified as "personal data", the system applies black masking (redaction) to the corresponding bounding box in the original label image, effectively concealing the sensitive information. This LLM-driven approach allows for context-aware redaction, going beyond simple keyword matching and enabling more accurate and nuanced personal data protection. The deterministic nature of the redaction decision, driven by the LLM's classification, ensures consistent and predictable system behavior, crucial for compliance-focused applications.

3.3. PDPA Compliance Integration

PDPA compliance is a central design principle of our system. By automating the redaction of personal data, we directly address the PDPA's requirements for data minimization and purpose limitation. The system is designed to redact only personal data, ensuring that downstream AI applications process anonymized label data, minimizing the risk of unintentional PDPA violations. The LLM's role in interpreting simplified PDPA guidelines represents a novel approach to embedding legal considerations directly into automated data processing workflows.

However, we recognize that AI interpretation of legal text is not perfect. Human oversight or review may be necessary in certain contexts to ensure full PDPA compliance, particularly in edge cases or ambiguous situations. Furthermore, the system's output

should be considered a tool to assist with compliance, not a guarantee of complete adherence to the PDPA. The ultimate responsibility for compliance rests with the organization using the system. The evaluation framework includes metrics assessing both accuracy and the potential for over-redaction. This reflects our commitment to balancing privacy protection with data utility, a key aspect of responsible PDPA compliance. We acknowledge that the simplified PDPA guidelines used in the LLM prompt may not cover all nuances of the full legal text, and future work will explore incorporating more comprehensive PDPA interpretations.

4. Evaluation and Experiments

To rigorously evaluate the performance of our AI-driven label redaction system, we conducted comprehensive experiments using a diverse and representative dataset of label images. Our evaluation focused on quantifying both the accuracy of text extraction achieved by the combined EasyOCR and Tesseract approach and the overall effectiveness of the system in accurately redacting personal data according to PDPA guidelines, utilizing a suite of established metrics.

4.1. Dataset Creation and Annotation

We assembled a dataset of 100 label images, carefully curated to reflect the diversity of real-world label data encountered in practical applications, specifically focusing on drug labels. The dataset comprised labels predominantly in the Thai language, reflecting our focus on PDPA compliance in Thailand, with a subset in English for comparison. The photographs were taken under varying lighting conditions and angles to simulate real-world capture scenarios.

To establish a robust ground truth for evaluation, each label image within the dataset underwent a manual annotation process. Trained human annotators, fluent in both Thai and English and possessing a thorough understanding of PDPA guidelines, performed three key annotation tasks:

1. Generating accurate ground truth text transcriptions for each label image, serving as the benchmark for OCR accuracy assessment.
2. Delineating precise bounding boxes around all text segments unequivocally classified as personal data according to PDPA criteria (names, addresses, hospital numbers, etc.), providing the ground truth for redaction effectiveness evaluation.
3. Categorizing each identified personal data segment into specific PDPA-relevant categories (e.g., names, addresses, phone numbers), enabling granular analysis of redaction performance across different PII types. This categorization also included a “non-PII” class for text that should not be redacted.

4.2. Experimental Setup

The experimental setup encompassed both software and hardware configurations designed to ensure reproducibility and facilitate rigorous evaluation. The AI-driven label redaction system was implemented using Python 3.9. It leveraged the EasyOCR library (version 1.7.1) [9] and the Tesseract library (version 5.3.3 via pytesseract wrapper) [10] for text extraction, and the ‘litellm’ library (version 1.16.14) [24] for simplified, provider-agnostic interaction with various LLMs. Image processing tasks were performed using OpenCV (version 4.9.0) [25]. Experiments were conducted on a dedicated workstation equipped with an NVIDIA GeForce RTX 4090 GPU with 24 GB of VRAM, providing sufficient computational resources for efficient OCR processing and LLM inference.

For redaction, we employed black masking, overlaying solid black rectangles on identified personal data bounding boxes to ensure unambiguous and visually clear concealment.

The evaluation procedure involved iterating through each label image in the annotated dataset. This included performing text extraction using both EasyOCR and Tesseract, combining their outputs, feeding the extracted text segments to the Qwen (Qwen/Qwen2.5-72B-Instruct-AWQ) LLM [8] for PDPA-aware classification, applying black masking based on the LLM's output, and subsequently calculating the predefined OCR performance and redaction effectiveness metrics by comparing the system's output against the created ground truth annotations. We also employed Gemini 1.5 Flash to provide explanations for a subset of the labels, assessing its ability to interpret the label content and identify potential PDPA violations in a different context (explanation rather than redaction).

5. Results

This section presents the results of our experiments, focusing on both the OCR performance and the overall redaction effectiveness of the AI-driven system.

5.1. OCR Performance Analysis

We analyzed the performance of the combined EasyOCR and Tesseract approach, focusing on its impact on the subsequent redaction task. While traditional OCR metrics like character error rate (CER) and word error rate (WER) are informative, they do not directly translate to the success or failure of PII redaction. A single character error in a name can be as detrimental as multiple errors in a non-PII word. Therefore, we qualitatively assessed the types of OCR errors and their potential consequences for PDPA compliance.

Our observations revealed several common error patterns: misrecognition of Thai characters (due to the script's complexity), incorrect word segmentation (due to the lack of spaces between words in Thai), errors in recognizing numbers and symbols, and partial text extraction (missing parts of words or phrases). Many of these errors were associated with low OCR confidence scores, suggesting that incorporating confidence thresholds in future work could improve performance. These OCR imperfections directly influence the LLM's ability to identify PII. A misspelled or fragmented name, phone number, or hospital number is less likely to be recognized as such by the LLM. This underscores the critical dependency between OCR quality and overall redaction performance. Our mitigation strategies included combining the outputs of both EasyOCR and Tesseract (prioritizing recall), using a targeted LLM prompt designed to handle minor OCR errors, and focusing on maximizing the useful information extracted for redaction, even if not perfectly accurate. We acknowledge that perfect OCR on real-world, noisy label images is a significant challenge.

5.2. Redaction Effectiveness

The core evaluation of our system focuses on its ability to accurately redact PII while minimizing over-redaction. We used four key metrics: redaction precision, redaction recall, F1-score for redaction effectiveness, and over-redaction rate. These metrics were calculated by comparing the system's automated redactions against human-annotated ground truth data, as detailed in Section 4. Table 1 presents the results.

Table 1. Redaction effectiveness metrics on label dataset (100 images).

Metric	Value
Redaction precision	92.5%
Redaction recall	83.2%
F1-score for redaction effectiveness	87.6%
Over-redaction rate	3.1%

The system achieved a high redaction precision of 92.5%. This indicates that when the system did redact a text segment, it was highly likely to be actual PII, minimizing false positives and preserving the informational content of the labels. The redaction recall of 83.2% shows that the system successfully identified and redacted a substantial majority of the PII present in the dataset. While commendable, this also indicates room for improvement in capturing all instances of PII, reducing false negatives.

The F1-score, which balances precision and recall, was 87.6%, demonstrating a robust and practically useful level of redaction effectiveness. The over-redaction rate was a low 3.1%, meaning that the system very rarely redacted non-PII text. This is crucial for maintaining the usability of the processed labels.

Further analysis of the results revealed that the primary cause of missed redactions (false negatives) was OCR errors, especially when dealing with stylized fonts or degraded image quality. Over-redactions (false positives), on the other hand, were typically caused by the LLM misinterpreting non-PII text that had superficial similarities to PII, such as a product code that resembled a phone number. These findings highlight the interconnectedness of the OCR and LLM components and point to areas for future refinement.

5.3. Qualitative Analysis with Gemini 1.5 Flash

To complement the quantitative evaluation of redaction effectiveness, we conducted a qualitative analysis using Gemini 1.5 Flash. This analysis aimed to assess the LLM's ability to understand the content of the drug labels after redaction, providing a different perspective on the system's performance. Instead of focusing on identifying PII for redaction, we tasked Gemini 1.5 Flash with explaining the purpose and usage instructions of the medication based on the redacted label image. This tests whether the redaction process preserves enough information for a separate LLM to understand the label's core meaning.

Figure 1 shows an example of a redacted label image. The text extracted by the combined EasyOCR and Tesseract from the original, unredacted image contained errors, as discussed in Section 5. The LLM-based redaction stage then identified and masked PII, including the hospital name, phone number, patient name, and hospital number.

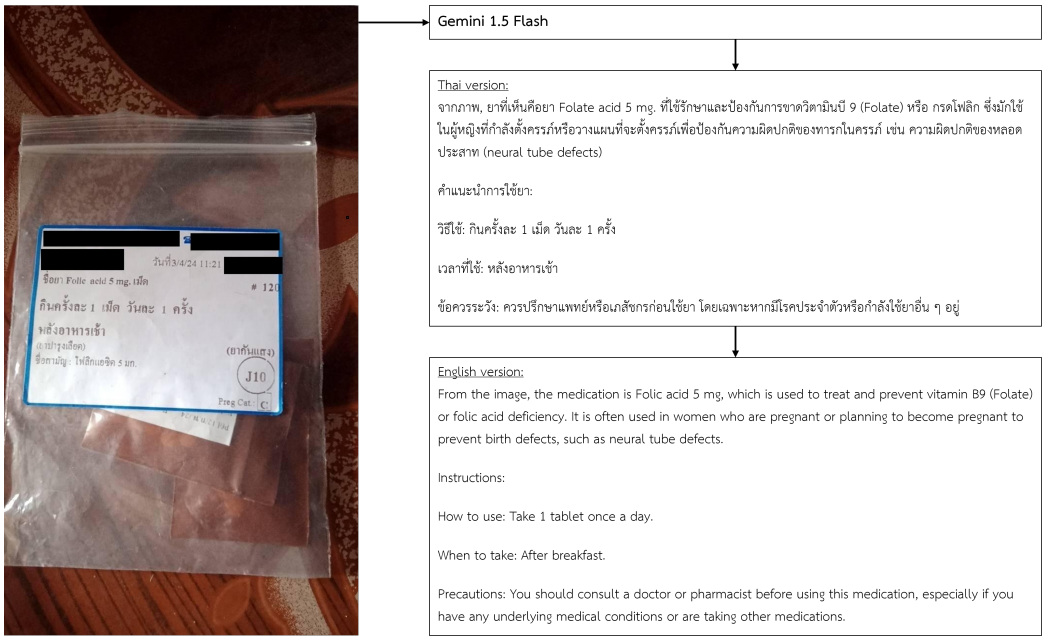

Figure 1. Example of a redacted drug label image. PII has been masked with black boxes.

We then provided the redacted image to Gemini 1.5 Flash, along with the following prompt (translated from the provided Thai prompt):

```
Explain the medication and instructions on this drug label.
```

Gemini 1.5 Flash generated the following response (translated from the provided Thai response):

```
From the image, the medication is Folic acid 5 mg, which is
used to treat and prevent vitamin B9 (Folate) or folic acid
deficiency. It is often used in women who are pregnant or
planning to become pregnant to prevent birth defects, such as
neural tube defects.

Instructions:

How to use: Take 1 tablet once a day.
When to take: After breakfast.
Precautions: You should consult a doctor or pharmacist before
using this medication, especially if you have any underlying
medical conditions or are taking other medications.
```

This result demonstrates that, despite the redaction of PII and the presence of OCR errors in the original text extraction, Gemini 1.5 Flash was able to correctly identify the medication (Folic acid), its purpose, and the dosage instructions from the redacted image. This indicates that the redaction process, while removing sensitive information, successfully preserved the essential information necessary for understanding the label's content. This qualitative analysis supports the quantitative findings, suggesting that the system achieves a good balance between privacy protection and data utility. It also demonstrates the potential of using a separate LLM for downstream tasks on redacted label data.

6. Discussion

This research investigated the feasibility and effectiveness of an AI-driven system using combined OCR and LLM techniques for automated personal data redaction on drug labels. The system specifically aimed for compliance with the Thai PDPA. It demonstrated a promising capability, achieving high redaction precision (92.5%) and reasonable recall (83.2%) coupled with a low over-redaction rate (3.1%). These results suggest potential for significantly reducing manual compliance efforts while preserving label utility. This was further supported by qualitative analysis where essential information remained understandable post-redaction.

The core strength lies in the system's automation and the LLM's context-aware redaction capabilities. These offer advantages over static rule-based methods by handling language variations and identifying PII semantically [11]. This adaptability is crucial for the complexities of real-world labels and PDPA nuances. Additionally, combining EasyOCR and Tesseract improved text extraction recall over using a single engine [9,10].

However, several limitations must be acknowledged. Firstly, the evaluation presented herein is based on a dataset of 100 label images, predominantly in Thai (Section 4.1). This dataset's size and limited diversity may not fully capture the extensive variability in real-world labels (different languages, formats, and qualities). Consequently, while sufficient for this initial feasibility study, the generalizability and robustness of the reported performance metrics require further validation on larger, more diverse datasets.

Secondly, the system's overall performance is intrinsically linked to the underlying OCR accuracy. Errors in text extraction, particularly with the complex Thai script or degraded image quality, were the primary cause of missed redactions (false negatives) [15]. This highlights the ongoing need for robust OCR advancements [12].

Thirdly, while the LLM showed strong capabilities using pre-trained models and prompt engineering, it occasionally misinterpreted non-PII (false positives). More critically, it relied on simplified PDPA guidelines provided via the prompt. This approach may not adequately capture the full complexity and potential ambiguities of the legal text. The system's ability to handle nuanced legal definitions and challenging edge cases (e.g., ambiguous text strings, partially redacted information potentially allowing re-identification) without domain-specific fine-tuning or explicit legal validation remains a significant limitation. This underscores the difficulty in fully replicating nuanced human legal judgment with current AI models alone. It highlights the critical need for validation by legal professionals and potentially human oversight in practical deployment scenarios, as discussed further in Section 7.

Finally, transitioning this system to a production environment requires careful consideration of practical aspects, particularly data security when interacting with external services. As noted in Section 3.2, our prototype utilized an external LLM API. While only text segments were transmitted, reliance on third-party services necessitates robust security measures beyond standard protocols for operational deployment under PDPA. This demands thorough vetting or alternative architectures like on-premise models, as discussed further in Section 7.

7. Conclusions and Future Work

This paper presented a novel AI-driven system combining OCR and LLM technologies for automated personal data redaction on drug labels to facilitate Thai PDPA compliance. The system achieved a promising balance between redaction accuracy, efficiency, and label utility preservation, demonstrating potential for practical application. This research contributes to the field of AI for legal compliance [22,23] by showcasing LLM application for data privacy regulation interpretation and highlighting the value of integrating multiple AI techniques.

Despite the promising results, significant avenues for future work are essential to enhance robustness, reliability, and compliance assurance:

1. Dataset Expansion for Robustness and Generalizability: Acknowledging the current dataset limitations (Section 6), a crucial next step is creating a significantly larger, more diverse benchmark dataset. This dataset should encompass varied label types, languages, layouts, and image qualities. This is essential for rigorous evaluation, improving generalizability, and enabling meaningful model fine-tuning.
2. Improving OCR Robustness: Further research should focus on enhancing OCR accuracy, particularly for Thai script and challenging label conditions. Exploring advanced OCR architectures, incorporating image pre-processing techniques (de-noising, de-warping), and fine-tuning OCR models on label-specific datasets could significantly improve performance [12].
3. Enhancing LLM Interpretation: Refining the LLM's understanding of the PDPA and its ability to distinguish PII from non-PII is crucial. This could involve experimenting with different prompting strategies, fine-tuning the LLM on a larger dataset of PDPA-relevant text, and incorporating legal domain knowledge [23]. Exploring different LLMs, including those specifically designed for legal tasks, could also be beneficial.
4. Validation with Legal Experts: Engaging legal and privacy professionals to rigorously validate the system's interpretation of PDPA requirements is essential. They should review its redaction decisions (especially in ambiguous situations) and its overall alignment with legal standards before any real-world deployment. This step is critical for ensuring genuine compliance and building trust.

5. Adaptive Redaction: Developing methods for adaptive redaction could further improve the balance between privacy and utility. Here, the level of redaction is adjusted based on the context and sensitivity of the information. This could involve incorporating confidence scores from both the OCR and LLM stages.
6. Human-in-the-Loop System: Integrating a human-in-the-loop component for review and validation of the system's redactions would enhance reliability. This is especially important for ambiguous cases or high-risk data and addresses the limitations of fully automated AI systems in legal compliance [3]. This could involve a user interface for reviewing and correcting the system's outputs.
7. Addressing API Security and Deployment Models: A critical consideration for practical deployment, particularly under strict regulations like the PDPA, involves the security implications of transmitting text data (even segments) to external LLM APIs, as employed in our prototype (Section 3.2). Future work should include rigorous security assessments of third-party API providers, potentially involving data processing agreements (DPAs). Furthermore, exploring and evaluating alternative deployment strategies, such as utilizing on-premise LLMs or models hosted within a secure private cloud infrastructure, is essential. This ensures greater control over the data environment and minimizes potential risks associated with external data processing, thereby reinforcing end-to-end PDPA compliance.
8. Ethical Considerations: A thorough ethical assessment of the broader implications of deploying such a system is necessary. This includes addressing potential biases in the AI models, ensuring transparency and accountability, and considering the impact on individual privacy rights.

By addressing these future research directions, AI-driven systems can become even more effective and reliable tools for ensuring data privacy compliance in an increasingly data-driven world.

Author Contributions: Conceptualization, P.T., B.S. and P.P.; methodology, P.T.; software, P.T.; validation, B.S. and P.P.; formal analysis, P.T., B.S. and P.P.; investigation, B.S. and P.P.; resources, P.T.; data curation, P.T.; writing—original draft preparation, P.T.; writing—review and editing, B.S. and P.P.; visualization, P.T.; supervision, B.S. and P.P.; project administration, P.T., B.S. and P.P.; funding acquisition, P.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets presented in this article are not readily available because they are private. Requests to access the datasets should be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Russell, S.J.; Norvig, P. *Artificial Intelligence: A Modern Approach*, 3rd ed.; Pearson: London, UK, 2016.
2. Voigt, P.; von dem Bussche, A. *The EU General Data Protection Regulation (GDPR): A Practical Guide*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017.
3. Goodman, B.; Flaxman, S. European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation". *AI Mag.* **2017**, *38*, 50–57. [CrossRef]
4. Ministry of Digital Economy and Society. Personal Data Protection Act B.E. 2562. 2019. Available online: https://mdes.go.th/law/detail/3577-Personal-Data-Protection-Act-B-E--2562--2019- (accessed on 14 February 2025).
5. Greenleaf, G. *Global Data Privacy Laws 2019: 132 National Laws & Many Bills*; Privacy Laws & Business International Report; Privacy Laws & Business: Pinner, UK, 2019; pp. 14–18.

6. Gartner Inc. Gartner Forecasts Worldwide Government IT Spending to Grow 8% in 2023. Press Release. Available online: https://www.gartner.com/en/newsroom/press-releases/2023-05-24-gartner-forecasts-worldwide-government-it-spending-to-grow-8-percent-in-2023 (accessed on 11 April 2025).
7. Ribarić, S.; Ariyaeeinia, A.; Pavešić, N. De-identification for Privacy Protection in Multimedia Content: A Survey. *Signal Process. Image Commun.* **2016**, *47*, 131–151. [CrossRef]
8. Yang, A.; Yang, B.; Zhang, B.; Hui, B.; Zheng, B.; Yu, B.; Li, C.; Liu, D.; Huang, F.; Wei, H.; et al. Qwen2.5 Technical Report. Version 1. *arXiv* **2024**, arXiv:2412.15115.
9. Jaided AI. EasyOCR. 2020. Available online: https://github.com/JaidedAI/EasyOCR (accessed on 11 April 2025).
10. Smith, R. An Overview of the Tesseract OCR Engine. In Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), Curitiba, Brazil, 23–26 September 2007; Volume 2, pp. 629–633.
11. Lison, P.; Pilán, I.; Sánchez, D.; Batet, M.; Øvrelid, L. Anonymisation Models for Text Data: State of the Art, Challenges and Future Directions. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Online, 1–6 August 2021; pp. 4188–4203.
12. Nachappa, C.H.; Rani, N.S.; Pati, P.B.; Gokulnath, M. Adaptive Dewarping of Severely Warped Camera-Captured Document Images Based on Document Map Generation. *Int. J. Doc. Anal. Recognit. (IJDAR)* **2023**, *26*, 149–169. [CrossRef] [PubMed]
13. Fan, X.; Zhao, W. Improving Robustness of License Plates Automatic Recognition in Natural Scenes. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 18845–18854. [CrossRef]
14. Nohlen, H.; Bakogianni, I.; Grammatikaki, E.; Ciriolo, E.; Pantazi, M.; Dias, J.; Salesse, F.; Moz Christofoletti, M.; Wollgast, J.; Bruns, H.; et al. *Front-of-Pack Nutrition Labelling Schemes: An Update of the Evidence*; JRC Technical Report JRC130125; Publications Office of the European Union: Luxembourg, 2022.
15. Chamchong, R.; Saisangchan, U.; Pawara, P. Thai Handwritten Recognition on BEST2019 Datasets Using Deep Learning. In Proceedings of the International Conference on Multi-Disciplinary Trends in Artificial Intelligence, Pattaya, Thailand, 12–16 November 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 152–163.
16. Agarwal, A.; Chattopadhyay, P.; Wang, L. Privacy Preservation through Facial De-identification with Simultaneous Emotion Preservation. *Signal Image Video Process.* **2021**, *15*, 951–958. [CrossRef]
17. Sweeney, L. K-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **2002**, *10*, 557–570. [CrossRef]
18. Arbelaez, A.; Climent, L. Iterative local search for preserving data privacy. *Appl. Intell.* **2025**, *55*, 1–14.
19. Dwork, C.; Roth, A. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.* **2014**, *9*, 211–407. [CrossRef]
20. Ghazi, B.; Golowich, N.; Kumar, R.; Manurangsi, P.; Zhang, C. Deep Learning with Label Differential Privacy. In Proceedings of the Advances in Neural Information Processing Systems 34 (NeurIPS 2021), Online, 6–14 December 2021; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2021; pp. 27131–27145.
21. Bonta, R. California Consumer Privacy Act (CCPA). State of California Department of Justice, Office of the Attorney General. 2022. Available online: https://oag.ca.gov/privacy/ccpa (accessed on 11 April 2025).
22. Walmsley, J. Artificial Intelligence and the Value of Transparency. *AI Soc.* **2021**, *36*, 585–595. [CrossRef]
23. Chalkidis, I.; Fergadiotis, M.; Malakasiotis, P.; Aletras, N.; Androutsopoulos, I. LEGAL-BERT: The Muppets Straight Out of Law School. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP, Online, 16–20 November 2020; pp. 2898–2904.
24. BerriAI. LiteLLM. 2023. Available online: https://github.com/BerriAI/litellm (accessed on 11 April 2025).
25. OpenCV Team. Open Source Computer Vision Library. 2023. Available online: https://opencv.org (accessed on 11 April 2025).