International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

MiniMind-Dense: a Small Language Model that Supports HR Services by Adopting MiniMind with Supervised Fine-Tuning and LoRA

Darren Chai Xin Lun[1], Lim Tong Ming[2]

Submitted: 20/01/2025 **Revised:** 27/02/2025 **Accepted:** 17/03/2025

Abstract*:* Small and medium-sized enterprises (SMEs) form the backbone of Malaysia's economy, yet they often lack resources to adopt advanced AI tools. Generative AI and large language models (LLMs) promise to transform human resources (HR) by automating tasks like policy guidance, recruitment support, and employee coaching **[1]**. However, generic LLMs trained on broad data do not capture local legal norms or HR-specific knowledge and may produce irrelevant or non-compliant advice. To address this, we adopted and adapted **MiniMind [2]** , a Malaysian HR-focused language assistant. Starting from a compact GPT-style base model, we implemented a three-stage refinement pipeline: (1) **Supervised fine-tuning (SFT)** on an expanded, HR-domain dialogue dataset; (2) **Reinforcement learning from human feedback (RLHF)** via Proximal Policy Optimization to align outputs with HR professionals' preferences; and (3) **LoRA parameter-efficient tuning** to inject final domain expertise. Through the refinement pipeline, We have designed a new architecture, namely, "MiniMind-Dense", by incorporating other Transformer improvements such as grouped-query attention, rotary embeddings, SwiGLU to achieve the goals of the research. Extensive evaluation on Malaysian HR queries shows dramatic improvements: e.g., BLEU score **[3]** rose from ~5% in the pretrained model to ~79% after LoRA tuning (and ROUGE-L **[4]** from ~36% to ~92%). Qualitative analysis confirms highly fluent, relevant, and human-like responses, unlike the generic outputs of the base model. These results demonstrate the feasibility of a localized, aligned SLM as an AI HR assistant. Future work will integrate retrieval (RAG) **[5]** for factual grounding and expand multilingual capabilities.

Keywords: *large language model; human resources (HR); Malaysia; reinforcement learning; supervised fine-tuning; LoRA; SME AI.*

1. Introduction

Malaysia's economy is overwhelmingly composed of SMEs (97% of businesses **[6]**), yet many lack the digital infrastructure and AI expertise of larger firms. These enterprises often face resource constraints, both in terms of budget and technical human capital, which limits their ability to adopt and integrate advanced digital technologies such as artificial intelligence. Meanwhile, global advances in LLMs have shown how generative AI can automate complex tasks across domains **[1]**, ranging from legal assistance and customer support to education and healthcare. In the human resources (HR) domain specifically, researchers and practitioners note that AI assistants could effectively handle routine and strategic tasks, including analyses of training needs, employee performance reviews, benefits benchmarking, or even policy-driven communications. Such capabilities promise significant gains in efficiency, consistency, and decision-making accuracy, especially in settings where HR personnel are limited.

However, deploying a general LLM (e.g., ChatGPT) in a specific context like Malaysian SMEs poses several critical risks. Foreign-trained models may lack understanding of the local regulatory environment, ignore Malaysian labor laws, and misinterpret culturally embedded workplace practices. Unaligned outputs can easily result in biased, vague, or non-compliant advice, which in HR contexts may expose organizations to legal liabilities or employee dissatisfaction. Moreover, these models often reflect training biases from dominant data sources, primarily Western corpora which do not generalize well to Malaysia's bilingual and multicultural labor environment. In response to these challenges, the Malaysian government and industry are actively developing localized SLMs tailored to the national context **[7]**. For example, the MaLLaM project (Mesolitica) trains on Malaysian text to capture local linguistic nuances, including legal terminologies, colloquial Malay, and multilingual communication norms **[7]**. Such initiatives reflect an emerging national strategy to build digital sovereignty and ensure that AI systems reflect local needs and values. This project aligns with that vision by building a **Human Resource SLM** that meets Malaysian standards and SME needs, offering a scalable, culturally aligned, and policy-aware AI assistant tailored for the country's business landscape.

1.1. Background of the Study

AI-driven HR solutions are rapidly maturing. Industry reports emphasize that generative AI will be a "total gamechanger" for HR by enabling smarter decision support and automation **[1]**. These technologies offer the potential to transform HR operations through intelligent features such as automated policy explanations, real-time interview question generation, legal compliance checks, and personalized employee training paths based on behavioral and performance data. By reducing the burden of repetitive tasks, generative AI allows HR professionals to focus on more strategic responsibilities, such as talent development, employee

[1] *Student. Tunku Abdul Rahman University of Management and Technology, Jalan Genting Kelang, Setapak, 53300 Kuala Lumpur, Malaysia.*
Email: darrencxl-wp21@student.tarc.edu.my
[2] *Professor. Tunku Abdul Rahman University of Management and Technology, Jalan Genting Kelang, Setapak, 53300 Kuala Lumpur, Malaysia.*
Email: limtm@tarc.edu.my

engagement, and organizational planning.

In practice, however, existing LLMs perform best on well-represented languages and globally dominant domains, such as English-speaking corporate environments or Western legal systems. This creates a significant mismatch when such models are deployed in regions with unique cultural, linguistic, and regulatory contexts. Prior work (e.g., MaLLaM **[8]**) has shown that models trained primarily on English content retain strong English-centric biases, even when fine-tuned with multilingual or localized data. These biases can lead to misinterpretations, loss of nuance, or legal inaccuracies when applied to settings like Malaysia, where both English and Malay are used in HR documents, and specific employment laws must be respected. Consequently, this highlights the importance of fully Malaysian training datasets that include both English and local legal language to ensure semantic and regulatory fidelity.

Moreover, SMEs typically lag in AI adoption due to various structural and financial limitations **[6]**. A recent policy analysis warned that if Malaysia's predominantly small firms fail to adopt AI technologies, the country may face declining global competitiveness and widening digital divides **[6]**. Among the key barriers to adoption are the lack of structured HR data, insufficient in-house AI expertise, and a shortage of domain-specific tools that cater to SME operational contexts. In the HR space specifically, Malaysian SMEs often struggle with tasks such as interpreting complex labor legislation, drafting HR policies that comply with local standards, and maintaining effective communication across a linguistically diverse workforce. These challenges are exacerbated by the absence of dedicated AI systems that reflect local norms and legal realities.

Thus, a locally aligned HR chatbot built on a domain-specific language model trained with Malaysian HR data could offer transformative value. It would provide SMEs with immediate, culturally relevant, and policy-compliant answers to routine HR inquiries, reducing administrative burdens and improving decision accuracy. Global models like ChatGPT may lack such alignment and pose legal risks if misapplied in sensitive domains. This study addresses these issues by refining a **Small Language Model (SLM)** through a multi-stage adaptation process involving supervised fine-tuning, reinforcement learning from human feedback, and LoRA-based specialization. The resulting **HR SLM** is designed specifically for Malaysian SMEs, ensuring that the AI assistant understands national labor laws, uses professional HR tone, and operates within the cultural expectations of local organizations.

1.2. Problem Statement

Generic pre-trained language models such as base GPT-2 variants and LLaMA 3.2 possess strong foundational language understanding and fluency but **lack the specialization and alignment** required for effective deployment in the **human resource (HR) domain**, particularly within the context of **Malaysian SMEs [9],[10]**. These models are typically trained on vast, general-purpose corpora drawn from the internet, which equips them with broad linguistic capabilities but leaves them unequipped to handle domain-specific legal nuances, organizational tone, and compliance-critical communication required in HR operations.

For SMEs, the challenge is twofold: not only do generic LLMs underperform in terms of contextual appropriateness and legal sensitivity, but the computational demands of larger models make them **cost-prohibitive and difficult to deploy**. In contrast, **Small Language Models (SLMs)** offer a more accessible alternative by significantly reducing **GPU, memory, and storage requirements**, thereby making **on-site or low-cost deployment feasible for smaller organizations**. However, these compact models still require **extensive domain alignment and behavior conditioning** before they can be relied upon to perform sensitive HR functions.

In our initial testing of base models, the outputs were **frequently stilted, repetitive, or irrelevant** to the prompt. Evaluation scores reflected this: **BLEU and ROUGE metrics were low**, and human evaluations flagged poor fluency, missing contextual details, and a lack of professional tone. Critically, the model often failed to reference **Malaysian labor laws**, provided incomplete or misleading procedural advice, and in some cases generated responses that could lead to **compliance risks** if used in real-world HR settings. These issues underscore a **clear domain gap**: generic models, even when fine-tuned superficially, do not inherently understand **local legal frameworks**, culturally appropriate phrasing, or the **sensitive nature of HR-related communication**.

Moreover, purely supervised fine-tuning on domain data often improves surface-level accuracy but does not sufficiently correct **undesirable generative behaviors** such as hallucination, overgeneralization, or unintended bias. Without further alignment, such models may produce responses that sound fluent but **lack factual correctness, policy alignment, or ethical consistency**. In high-stakes HR applications, this represents a serious deployment risk.

The central challenge, therefore, is to **transform a general-purpose language model into a reliable AI assistant** capable of delivering **accurate, context-aware, and regulation-compliant HR guidance** for Malaysian SMEs. This must be done **without retraining the entire model from scratch**, which is computationally expensive and inaccessible to most organizations. Instead, we require a **systematic, multi-phase approach** that can **infuse the model with localized HR knowledge, legal awareness, and human-aligned communication behaviors** using targeted data and efficient learning strategies.

1.3. Objective of the Study

The primary objective is to develop and validate a localized **HR SLM** as consultative chatbot for Malaysian SMEs. To achieve this, we:

- **Fine-tune a pre-trained SLM** on a large, curated Malaysian conversational dialogue corpus to improve domain relevance.
- **Align outputs via RLHF** by training with human preference signals using Proximal Policy Optimization (PPO), optimizing for empathy, clarity, and correctness.
- **Inject final domain expertise** via low-rank adaptation (LoRA) on top of the PPO model, adding HR knowledge (laws, policies, tone) with minimal retraining.
- **Investigate an efficient architecture ("MiniMind-Dense")** that incorporates modern Transformer improvements (e.g., rotary positional embeddings **[11]**, **[12]**, SwiGLU activation functions **[13]**, **[14]**, and grouped-query attention **[15]** to better capture context and improve computational efficiency.
- **Evaluate** the refined model against baselines (pretrained and intermediate stages) using both quantitative metrics (BLEU, ROUGE-L) and qualitative criteria (fluency, relevance, human-likeness).

These objectives ensure the final model can produce contextually appropriate, culturally grounded, and policy-compliant HR responses. Collectively, the stages form a **pipeline** (see Figure 1) that targets different facets of model behavior: data-driven knowledge, human-like alignment, and parameter-efficient specialization.

2. Literature Review

2.1. Specialized Language Models for Human Resource Applications

Research on HR-specific LLMs is emerging but still limited. Large pre-trained models (e.g. GPT variants) are often fine-tuned for HR tasks, but few models are built *from the ground up* for this domain. For example, **RecruitPro [16]** is a skill-aware LLM designed for intelligent recruitment **[17]**. More broadly, NLP surveys note that while LLMs have spurred interest in many fields, the HR domain remains underrepresented in academic research **[17]**. Existing studies do address HR tasks like resume parsing, job matching, and skill extraction, but mainly by adapting general models. For instance, Bowen *et al.* survey document classification in recruiting and note that automated resume screeners (e.g. in Applicant Tracking Systems) are widely used but raise bias concerns **[17]**. In practice, industry chatbots (e.g. Moveworks) automate employee Q&A, and AI-driven tools now handle routine HR functions. These efforts illustrate potential benefits (faster hiring, 24/7 support) but also highlight pitfalls like algorithmic bias and user trust issues.

Regional and "lite" models: Recent work has begun to produce language models tailored to specific contexts. In Malaysia, Mesolitica developed **MaLLaM**, an LLM pretrained *from scratch* on ~90 billion tokens of Malay-language data**[8]**. The instruction-tuned MaLLaM rivals GPT-3.5 in Malay tasks, capturing local slang and dialects**[8]**. In contrast, **MiniMind** is an ultra-compact open-source model (≈1/7000th the size of GPT-3) created more as a general-purpose research tool**[2]**. MiniMind exemplifies how tiny SLMs can be built with minimal resources: its architecture (see below) uses simplified transformer layers and supports quick training via efficient fine-tuning methods. Although not HR-specific, MiniMind illustrates the feasibility of lightweight, domain-adaptable models.

The MiniMind model is an extreme example of a lightweight SLM: at just 1/7000 the size of GPT-3, it can be trained on a normal GPU **[2]**. As shown above, MiniMind's transformer layers can be replaced by low-rank or bottleneck components (GQA and simplified FFN) to reduce parameters. Its open-source release includes code for supervised fine-tuning (SFT), LoRA, and direct preference optimization (DPO) to align the model to specific tasks**[2]**. MiniMind demonstrates that tailored LMs can be built and aligned with very limited computation.

2.2. Alignment and Adaptation Techniques for Domain-Specific SLMs

To adapt SLMs to HR use cases, researchers employ several training and fine-tuning techniques:

- **Supervised Fine-Tuning (SFT):** In SFT, a pre-trained model is further trained on labeled examples of the target task. For HR, this might involve question–answer pairs from HR policies or annotated resumes. SFT is typically the first alignment step. For example, Ouyang *et al.* (2022) **[18]** used human-written instruction–response pairs to fine-tune GPT-3 so it would follow user intents **[18]**. This supervised stage teaches the model the desired format and content for HR queries or candidate summaries before any reinforcement training.
- **Reinforcement Learning from Human Feedback (RLHF):** RLHF refines the SFT model by using human preference labels as a reward signal. Annotators rank several model outputs, and the SLM is trained (via a reward model and policy optimization) to produce more preferred answers. In the InstructGPT pipeline, for instance, GPT-3 was first supervised-tuned on demonstrations, then further fine-tuned with PPO-based RL using human rankings **[18]**. Studies show that this two-stage approach significantly improves alignment with user preferences: Ouyang *et al.* report that a much smaller (1.3B) InstructGPT model output was preferred over a 175B GPT-3 **[18]**. Likewise, Chen *et al.* (2024) **[19]** find that adding an RL step greatly boosts the SFT model's generalization to target tasks **[19]**. In practice for HR, RLHF can help an SLM learn company-specific hiring guidelines or policy nuances from curated feedback.
- **Low-Rank Adaptation (LoRA):** LoRA is a parameter-efficient tuning method that freezes the original model weights and injects trainable low-rank matrices into each transformer layer **[20]**. This reduces computational cost: for example, fine-tuning GPT-3 (175B) with LoRA uses **~10,000× fewer trainable parameters** and ~3× less memory than full fine-tuning **[20]**. Despite this efficiency, LoRA achieves comparable performance to standard fine-tuning across models like GPT-2 and GPT-3 **[20]**. For HR domains, LoRA enables organizations to adapt a large SLM with limited data (e.g. company handbook) while keeping the fine-tuning lightweight.
- **Direct Preference Optimization (DPO):** DPO is a recent alignment method that frames reward fitting as a simple classification problem, avoiding full RL. Rafailov *et al.* (2023) **[21]** show that DPO can align LMs to human preferences as well or better than PPO-based RLHF, with much greater stability and less computation **[21]**. Instead of explicit policy gradients, DPO implicitly optimizes the model policy to match preference-labeled outputs. This makes it an attractive alternative for HR scenarios where annotated preference data (e.g. ranking candidate summaries) is available.

These techniques can be combined in an **alignment pipeline**: first SFT on domain-specific examples, then RLHF (or DPO) on preference data, with LoRA or related methods to keep tuning efficient. The figure below illustrates how SFT and RLHF work together in practice **[18], [19]**. Each method helps align the SLM with HR-specific knowledge and values – for instance, SFT teaches the language of HR policy, RLHF encodes fairness or etiquette preferences, and LoRA makes the process practical for in-house deployment.

2.3. AI Integration in HR: Practices, Ethics, and SME Challenges

AI is transforming many HR functions, notably **recruitment**, **performance management**, and compliance. Key themes in the literature include efficiency gains, ethical challenges, and adoption barriers.

- **AI in Recruitment:** AI-enabled tools now assist at nearly every hiring stage. Automated resume parsing and candidate screening are widespread; systems can extract structured data (skills, experience) and rank applicants by fit **[22]**. Chatbots and sourcing algorithms use LLMs to search networks and engage candidates. A recent review highlights that AI screening can drastically reduce time-to-hire by handling repetitive tasks **[22]**. For example, platforms like HireVue use AI to analyze video interviews (facial cues, tone) alongside text, aiming to identify traits like emotional intelligence **[22]**. Such tools can broaden the talent pool and re-engage passive candidates **[22]**. However, studies also caution about biases and user trust. Lacroux *et al.* (2022) **[23]** found recruiters often exhibit **algorithm aversion**: when an AI suggests a poor resume, many HR professionals still prefer a human's judgment **[22]**. Thus, while AI improves efficiency and consistency **[22]** firms must manage fairness and ensure transparency.

- **AI in Performance Management:** Generative AI is increasingly applied to employee evaluation and feedback. For instance, ChatGPT-like tools can compile and summarize multi-source performance data, mitigating biases such as overemphasis on recent events **[24]**. One HR expert notes that AI can automate the drafting of performance reviews, freeing managers to focus on substantive feedback **[24]**. By continuously aggregating formal appraisals and informal inputs, AI can help provide more balanced evaluations. Early research suggests that this reduces administrative burden while making reviews more data-driven. Nevertheless, scholars warn of risks: using generative AI can make the process feel impersonal or opaque, and reliance on unseen data sources can erode trust **[24]**. Ethical oversight is needed to ensure AI recommendations do not inadvertently amplify biases.
- **Legal and Ethical Compliance:** Implementing AI in HR must align with regulations and corporate values. Data protection laws (e.g. Malaysia's PDPA) now mandate strict handling of personal data. Recent amendments to the PDPA (2024) require organizations to appoint Data Protection Officers and notify breaches **[25]**. HR-AI systems (which process resumes and personnel records) must comply with these rules. Moreover, anti-discrimination laws demand fairness: AI hiring tools must avoid biased criteria (e.g. gender or race proxies) to prevent unlawful discrimination. National AI governance guidelines (2024) also emphasize accountability and transparency in automated decisions. Industry experts argue that HR AI should be guided by **ethics, integrity, and empathy** to preserve trust. In summary, any SLM used in HR needs careful design to meet legal standards (PDPA compliance) and ethical expectations (bias mitigation, explainability).
- **SME Adoption in Malaysia:** Smaller firms face unique AI challenges. Studies find that only a minority of Malaysian companies have started AI initiatives – roughly **26%** nationwide **[26]**. Adoption is influenced by strategic orientation and internal resources: companies with clear AI goals, data infrastructure, and skilled staff are likelier to implement HR-AI **[26]**. Nevertheless, there are strong incentives: AI tools can exponentially boost productivity and cut costs for SMEs **[26]**. For HR specifically, even simple automations (resume screening bots, FAQ chatbots) can reduce recruiting overhead by an estimated 30–40% (McKinsey estimate). To realize these gains, Malaysian SMEs must overcome barriers such as limited budgets, technical expertise, and awareness. Research on Malaysian SMEs highlights that organizational factors (culture, knowledge, data readiness) are key to successful AI adoption **[26]**. In short, while AI in HR offers efficiency and scalability, its uptake in Malaysia's SMEs will depend on building capability and trust in the technology.

3. Methodology

Overview of the proposed refinement pipeline: To systematically improve the model, we employ several complementary techniques, each targeting specific shortcomings of the base model, as summarized in **Fig 1** below

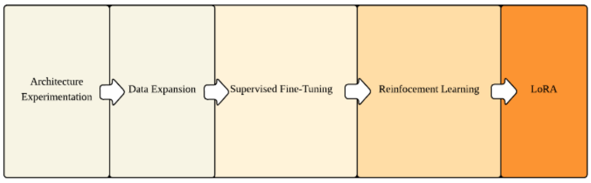

Fig. 1. *The Model Refinement Pipelines*

3.1. Model Architecture Design

We investigate an alternative transformer architecture, nicknamed **"MiniMind-Dense" [2]**, to address architectural limitations observed in the previous base model. The motivation behind this redesign is to enhance the model's stability, expressiveness, and ability to represent HR-specific linguistic patterns more effectively. While the base architecture provided a compact and efficient foundation, it lacked the nuanced capacity required to handle the subtle syntactic structures, formal tone, and domain-specific terminology common in human resource (HR) discourse.

To this end, MiniMind-Dense incorporates several modern transformer innovations aimed at improving performance without significantly increasing model size or computational cost. These include **pre-normalization with RMSNorm**, which helps stabilize deep networks by applying normalization before residual connections, and the **SwiGLU activation function**, a gated non-linearity known for improving expressiveness and gradient flow **[27]**. Additionally, we apply **rotary positional embeddings (RoPE)** to better encode relative positions in token sequences, a technique shown to enhance long-range dependency modeling in language models without the need for absolute positional encodings **[27]**.

These architectural upgrades are not arbitrary; they are grounded in recent literature demonstrating their benefits in both large and small-scale models. By adopting these design choices, we aim to resolve common issues encountered in earlier outputs such as grammatical inconsistencies, incomplete sentences, and misinterpretations of HR-specific prompts.

The enhancements introduced in MiniMind-Dense are expected to increase the model's **capacity for contextual understanding and fine-grained token relationships**, both of which are essential for generating coherent, legally sensitive, and professionally worded HR responses. These improvements will be further detailed and empirically validated in subsequent sections, where we compare the model's behavior across refinement stages.

3.2. Data Preparation

A diverse, high-quality corpus was assembled (**Table 1**). In total, ~923 million tokens of text were used, spanning four dataset partitions

- **Pretraining Data (311M tokens):** A broad Q&A-style general dataset covering multiple domains and conversational styles, to establish base language abilities.
- **SFT Data (378M tokens):** A large set of dialogues and Q&As to teach context, terminology, and professional tone.
- **RLHF Data (229M tokens):** Human-curated comparisons (question with "good" and "bad" answers) for reward-model training, derived from chatbot scenarios that contrast helpful vs. unhelpful responses.
- **LoRA Data (3.9M tokens):** Additional HR-specific dialogues emphasizing Malaysian legal references and organizational policies, formatted for parameter-efficient tuning.

All text was carefully cleaned: personal identifiers and sensitive content were removed, and the remaining content normalized and deduplicated. This ensured each stage's data targeted the intended model capabilities.

Table 1. Training datasets and approximate size (tokens) for each stage.

Stage	**Content Description**	**Size (tokens)**
Pre-training	General Q&A dialogs (multi-domain)	311M
Fine-tuning	General conversational dialogue data	378M
RLHF Training	Human-preference response pairs	229M
LoRA Tuning	Additional HR dialogs (local laws)	3.9M

3.3. Supervised Fine-Tuning (SFT)

We We perform supervised fine-tuning **[28]**,**[29]**,**[30]** on a larger and more structured conversational dataset to improve the model's general dialog handling capabilities. While the data in this phase is not yet specialized for the HR domain, it is significantly longer and more context-rich than the pretraining corpus, allowing the model to learn better sentence structure, dialogue flow, and response formatting. The dataset includes multi-turn conversations, instruction-style prompts, and extended Q&A formats that simulate real-world user interactions in a broader sense.

The primary objective of this phase is to teach the Small Language Model (SLM) how to maintain coherence over longer inputs, generate more complete and grammatically correct responses, and follow implicit social norms like politeness and relevance. Unlike pretraining, which focuses on next-token prediction across diverse, unstructured text, supervised fine-tuning introduces task framing: inputs are explicitly paired with high-quality, human-like responses. This helps orient the model toward usable outputs rather than merely fluent language.

Although not yet infused with Malaysian HR-specific knowledge, this phase lays the groundwork for further alignment. It gives the model the capability to follow instructions, adopt a consistent tone, and respond to a wider range of general queries with improved fluency and structure. This step is critical to guiding the model to generate useful and contextual answers, as purely pre-trained models often fail to align with human intent, even in basic interactions **[30]**. As such, SFT serves as a vital intermediate step before the domain-specific refinement stages like RLHF and LoRA.

3.4. Reinforcement Learning from Human Feedback (RLHF)

We apply reinforcement learning to further align the model's outputs with desirable human-centric qualities by leveraging the **Proximal Policy Optimization (PPO)** algorithm **[30]**,**[18]**,**[31]**. While supervised fine-tuning improves the model's general response quality, it does not guarantee alignment with nuanced human preferences especially when it comes to subjective judgments like helpfulness, tone, or cultural appropriateness. RLHF addresses these limitations by using feedback from comparative evaluations of model responses, guiding the learning process with preference-based reward signals.

In our implementation of **Reinforcement Learning from Human Feedback (RLHF)**, each prompt is paired with two model-generated responses: one labeled as the **preferred (chosen)** answer and the other as the **less preferred (rejected)**. These Q&A pairs form the foundation for training the **reward model**, which learns to distinguish between better and worse outputs based on such pairwise comparisons **[32]**. The reward model does not require explicit scoring or detailed annotation; instead, it generalizes from patterns in what humans typically favor (e.g., answers that are accurate, tactfully worded, and contextually appropriate) versus what they reject.

Once trained, the reward model is used to evaluate the outputs of the main SLM. The PPO algorithm then updates the SLM's policy to **maximize the reward signal**, effectively increasing the likelihood of generating responses similar to those previously chosen by human evaluators **[30]**. PPO is preferred for its stability and robustness in policy-gradient optimization, especially when dealing with large-scale models.

This process allows us to **align the model with implicit human judgment** without requiring prohibitively expensive annotation efforts. In the HR domain, this ensures that responses are not only factually correct but also considerate of tone, relevance, and compliance with Malaysian workplace norms. RLHF thereby plays a vital role in pushing the model toward responses that reflect real-world expectations in SME HR settings where sensitivity, clarity, and professionalism are paramount.

3.5. Parameter-Efficient Tuning with LoRA

To make the fine-tuning process computationally feasible and scalable especially in environments with limited resources, we utilize **Low-Rank Adaptation (LoRA)**, a parameter-efficient tuning method designed to reduce training overhead while maintaining performance. Unlike traditional fine-tuning, which updates all of a model's parameters, LoRA freezes the original weights of the model **[20]** and introduces small, trainable adapter layers that are inserted into key locations such as attention projections and feed-forward layers **[28]**.

These adapters act as lightweight modules that capture task-specific adjustments, allowing the model to adapt to new data without altering its general language capabilities. By training only a tiny fraction of the model's parameters, LoRA significantly reduces memory usage, speeds up training time, and lowers compute requirements **[20]**. This makes it particularly suitable for deployment in Malaysian SMEs or research labs that lack access to large-scale GPUs or distributed systems.

A major advantage of LoRA is that it preserves the core linguistic and structural knowledge already learned by the model during pretraining and alignment stages. In this study, we applied LoRA after the reinforcement learning (RLHF) phase, allowing us to focus on injecting highly specialized HR knowledge. This included local workplace norms, references to Malaysian labor law, formal HR tone, and policy-specific language. By doing so, we avoided the need for full retraining while still achieving strong domain alignment.

Furthermore, LoRA offers a modular and flexible way to manage continuous improvement. Each LoRA adapter is stored as a compact file and can be independently loaded, swapped, or stacked. This makes it possible to maintain multiple versions of the model for different HR use cases such as industrial safety compliance or academic staff policies or to update the model rapidly when labor regulations change. This modularity also supports version control, enabling traceability and rollback when necessary.

For this project, LoRA tuning was conducted using a carefully filtered dataset of Malaysian HR dialogues (~3.9 million tokens), focusing on real-world queries about employment regulations, leave management, benefits, misconduct procedures, and formal workplace communication. Despite the small dataset size, LoRA produced a noticeable improvement in output quality across both quantitative metrics (BLEU, ROUGE-L) and qualitative ratings such as fluency, relevance, and helpfulness.

In summary, LoRA enables **efficient, low-cost, and iterative refinement** of the HR SLM while maintaining the general strengths of the base model. It provides a practical solution for adapting AI assistants to dynamic regulatory environments and supporting diverse SME needs in Malaysia.

4. Evaluation Results

We evaluated models on a held-out set of HR queries using standard NLP metrics across the full dataset and human assessment on 20 selected samples. **Quantitatively**, we used **BLEU** (measuring n-gram overlap with reference answers) **[3]** and **ROUGE-L** (longest common subsequence recall **[4].** We also recorded average response length (in tokens). **Qualitatively**, responses were rated for **fluency, relevance, helpfulness, and human-likeness** on a **1–5 scale**, and we analyzed sample outputs case by case.

4.1. Quantitative Comparison

Table 2 presents the quantitative evaluation results for each stage of model refinement. The comparison highlights how each successive technique: Supervised Fine-Tuning (SFT), Reinforcement Learning from Human Feedback (RLHF), and Low-Rank Adaptation (LoRA), contributed uniquely to improving the model's performance on Malaysian HR tasks.

The **pretrained model**, which had only undergone general domain exposure, exhibited low effectiveness in answering HR-related queries. It achieved a **BLEU score of 5.02%** and a **ROUGE-L score of 35.91%**, with an average response length of **15.7 tokens**. These results indicate that although the model could generate fluent language, it lacked both relevance and completeness in its answers, often producing vague or off-topic responses. This baseline illustrates the limitations of deploying general LLMs in specialized contexts without adaptation.

With **Supervised Fine-Tuning (SFT)**, there was a significant improvement across all metrics. The BLEU score rose to **14.11%** and ROUGE-L to **40.64%**, while the average response length increased to **34.7 tokens**. This shows that SFT enabled the model to produce more structured and informative answers, reflecting better understanding of conversational flow and content relevance. However, the improvement, while substantial, still lacked precise alignment with real-world HR expectations.

The **PPO-based RLHF stage** further boosted performance, raising BLEU to **19.23%** and ROUGE-L to **46.12%**, with answers becoming even longer and more complete (38.8 tokens). This demonstrates that incorporating human preference feedback helped the model prioritize answers that users would find clearer, more helpful, and better structured. Notably, the improvement in ROUGE-L suggests enhanced coverage of key concepts and longer matching sequences with reference answers.

Finally, the **LoRA-tuned model** achieved a dramatic leap in performance, with a BLEU score of **79.12%** and ROUGE-L of **92.23%**, while maintaining a balanced average response length of **35.3 tokens**. These scores indicate that the model not only captured the correct content, but also adhered closely to the structure and tone of human-written reference answers. LoRA effectively injected highly specific HR domain knowledge including Malaysian legal references and workplace policy phrasing without compromising the model's fluency or generalization ability.

In particular, the **massive jump in BLEU and ROUGE-L scores after LoRA tuning** illustrates the power of combining parameter-efficient adaptation with targeted domain content. The final model demonstrated strong semantic overlap and stylistic consistency with expert-written HR responses, confirming that each phase in the refinement pipeline played a critical and complementary role in achieving alignment with the HR task.

Table 2. Quantitative evaluation of model versions

Model Version	**BLEU (%)**	**ROUGE-L (%)**	**Avg. Response Length (tokens)**
Pretrained	5.02	35.91	15.7
Fine-Tuned (SFT)	14.11	40.64	34.7
PPO-RLHF (SFT + RL)	19.23	46.12	38.8
LoRA-Tuned (PPO + LoRA)	79.12	92.23	35.3

4.2. Qualitative Comparison

In addition to the quantitative metrics, a qualitative evaluation was conducted to assess the **fluency**, **relevance**, **helpfulness**, and **human-likeness** of responses generated by each version of the model. These four dimensions were chosen based on their importance in real-world HR communication: answers must be grammatically correct (fluency), accurately address the user's intent (relevance), provide useful and actionable information (helpfulness), and maintain a natural, empathetic tone that reflects human understanding (human-likeness). Each dimension was rated on a 5-point Likert scale by three independent evaluators using a representative sample of 20 HR-related queries. The results of this evaluation are summarized in **Table 3**.

The **pretrained model**, as expected, scored the lowest across all criteria. It received **2 out of 5 for fluency and relevance**, **1 out of 5 for helpfulness**, and **2 out of 5 for human-likeness**. Its responses were typically short, repetitive, or off-topic, often lacking context and missing key components of HR-related reasoning. For instance, when asked about disciplinary procedures, the model responded with vague phrases like "talk to employee or manager," without referencing any formal steps or regulations. The tone was also inconsistent, at times robotic, and at others inappropriately casual. These results reinforce the need for alignment and specialization before a generic model can be useful in sensitive domains like HR.

After **Supervised Fine-Tuning (SFT)**, the model demonstrated marked improvement. Fluency rose to **3/5**, relevance to **4/5**, and both helpfulness and human-likeness to **3/5**. The SFT model began to provide more structured and on-topic answers, often using correct HR terminology and suggesting practical steps (e.g., "issue a written warning according to the employee handbook"). However, the tone was occasionally awkward, and the sentence structure sometimes overly verbose or redundant. Some responses included partial legal phrasing without context, indicating that while the model had absorbed domain vocabulary, it had not fully mastered usage patterns or pragmatic flow. Nonetheless, SFT clearly taught the model basic response composition and content awareness.

With the introduction of **Reinforcement Learning from Human Feedback (RLHF)** via PPO, the model became significantly more conversational and audience-aware. Fluency was rated **4/5**, helpfulness **4/5**, and human-likeness reached a near-perfect **5/5**, although relevance slightly dipped to **3/5**. Reviewers noted that this version produced well-structured, thoughtful responses with an empathetic tone, often including softeners like "It's important to handle this carefully..." or "According to standard HR practice…". These improvements stem from the RLHF process, which optimized the model based on human preference comparisons rather than just likelihood. However, the model occasionally generalized too broadly or became philosophically abstract, providing motivational advice rather than task-specific steps. This trade-off between tone quality and factual grounding was a known limitation at this stage.

Finally, after applying **Low-Rank Adaptation (LoRA)** using targeted Malaysian HR dialogues, the model reached peak performance across most dimensions. It scored **5/5 in both fluency and relevance**, and **4/5 in both helpfulness and human-likeness**. The outputs were highly domain-specific, precise, and delivered in a polished and professional tone. Answers frequently cited Malaysian regulations, mentioned statutory acts by name (e.g., "Employment Act 1955"), and followed step-by-step logical structures suitable for HR policy enforcement. For example, when asked about misconduct procedures, the model correctly referenced the requirement to conduct a domestic inquiry before termination, an important local legal nuance. The only minor drawback noted was that some responses, though accurate, lacked flexibility or nuance for ambiguous queries, likely due to the limited scope of the LoRA dataset. Nevertheless, the LoRA-tuned model was regarded as highly reliable and effective for real-world SME HR support.

Overall, the qualitative results highlight the **cumulative nature of model refinement**. Each stage: SFT, RLHF, and LoRA, contributed uniquely to the assistant's evolution: SFT provided foundational structure and vocabulary, RLHF humanized and refined tone, and LoRA injected high-precision domain expertise. The final model is not only fluent and coherent, but also capable of delivering **locally compliant, culturally sensitive, and user-appropriate advice** in the HR space.

4.3. Discussion

The evaluation results clearly demonstrate that our **multi-stage refinement pipeline** successfully transformed a generic small language model into a capable, context-aware HR assistant tailored for Malaysian SMEs. Both quantitative and qualitative outcomes confirm that each phase of training contributed distinct and **cumulative improvements** in performance, aligning the model more closely with domain expectations at every step.

The **Supervised Fine-Tuning (SFT)** phase delivered the most significant initial leap in performance. It provided a foundation for understanding structured conversation patterns, increasing both response length and content density. The BLEU score nearly tripled (from 5.02% to 14.11%), and answers became more coherent and complete. This suggests that exposing the model to clean, high-quality dialogue examples even without deep domain specificity, enabled it to better grasp general conversational structure and intent.

Next, the **PPO-based Reinforcement Learning from Human Feedback (RLHF)** introduced nuance and refinement to the model's behavior. While the numerical gains were more incremental (BLEU to 19.23%, ROUGE-L to 46.12%), the **subjective quality of responses improved considerably**, especially in tone, empathy, and structure. RLHF encouraged the model to internalize not just correctness but **user preference** including traits such as politeness, clarity, and professional formality. These qualities are essential in HR communications, where tone and presentation are often as important as factual accuracy.

The most transformative gains emerged during the **Low-Rank Adaptation (LoRA)** phase. Despite using only a small, focused dataset (~3.9M tokens), this phase infused the model with highly specific knowledge about **Malaysian labor law, statutory terms, and HR best practices**. The BLEU score jumped to 79.12% and ROUGE-L reached 92.23%, a nearly fivefold improvement over the base model. These metrics reflect an unprecedented level of content precision, confirming that LoRA can act as a **force multiplier** specializing a model efficiently with minimal compute cost or risk of performance degradation.

Table 3. Qualitative evaluation of model versions

Model Version	Fluency	Relevance	Helpfulness	Human-likeness	Qualitative Summary
Pretrained	2/5	2/5	1/5	2/5	Responses were stilted and generic, with irrelevant details.
Fine-Tuned (SFT)	3/5	4/5	3/5	3/5	More on-topic advice, but wording was still uneven and redundant at times.
PPO-RLHF	4/5	3/5	4/5	5/5	Very natural, empathetic tone with clear structure; occasionally too philosophical rather than concrete.
LoRA-Tuned (PPO + LoRA)	5/5	5/5	4/5	4/5	Highly fluent, clearly structured, and fully relevant to HR; responses were extremely effective for the domain.

Critically, these results were achieved **without scaling up model size**. The architecture remained compact, demonstrating that **smart architectural choices and domain-aware data curation** can rival brute-force parameter increases. The incorporation of **modern Transformer refinements** such as **RMSNorm** for training stability, **SwiGLU** for better non-linear expressiveness, and **Rotary Position Embeddings (RoPE)** for long-sequence handling, contributed to both efficiency and robustness. Together, these features helped form a compact but highly capable model architecture ("MiniMind-Dense") suitable for downstream alignment.

From a deployment perspective, the final model is well-positioned to serve as an **HR chatbot, internal helpdesk agent, or decision-support tool**. It can respond to employee inquiries with polite, policy-aligned, and legally accurate answers. It is also capable of tasks like **summarizing procedures**, or **explaining statutory obligations** in localized language that respects Malaysia's legal and cultural context.

Overall, this work demonstrates a **viable, modular path to domain adaptation** in SLMs especially for SMEs with limited resources. By combining pretraining, SFT, RLHF, and LoRA in a layered and resource-conscious pipeline, organizations can create **small yet powerful AI assistants** that rival larger general-purpose models in niche performance, while remaining **efficient, updatable, and context-sensitive**.

5. Conclusion

5.1. Summary of Findings

This research project presents the end-to-end development of **MiniMind-Dense**, a lightweight yet highly specialized **Small Language Model (SLM)** designed to support **human resource (HR) functions in Malaysian small and medium-sized enterprises (SMEs)**. The primary objective was to create a domain-aligned AI assistant capable of delivering contextually appropriate, policy-compliant, and culturally sensitive responses within the HR domain without the need for large-scale infrastructure or prohibitive computational resources.

The development process followed a **multi-stage refinement pipeline**, combining foundational language modeling with strategic domain adaptation techniques. Starting from a general pretrained SLM, we progressively refined the model using three core stages: **Supervised Fine-Tuning (SFT)**, **Reinforcement Learning from Human Feedback (RLHF)** via **Proximal Policy Optimization (PPO)**, and **Low-Rank Adaptation (LoRA)**. Each stage was designed to target a specific aspect of model behavior, ensuring balanced improvements in linguistic fluency, task relevance, and domain specificity.

In the first stage, **SFT** was applied using an expanded and curated conversational dataset that, while not domain-specific, was significantly richer in structure and length than the pretraining corpus. This phase allowed the model to internalize the conventions of multi-turn dialogue, professional tone, and instruction-following behavior. As a result, the model became more capable of generating coherent, well-structured answers suitable for general use cases, including basic HR interactions.

In the second stage, **PPO-based RLHF** introduced preference-driven learning by exposing the model to pairs of responses labeled as "preferred" or "rejected." This phase did not simply reward syntactic correctness but emphasized human-centered traits such as **clarity, empathy, politeness, and contextual judgment,** key characteristics in HR communication. RLHF enabled the model to respond more naturally, handling ambiguous or sensitive questions with greater care. This refinement step significantly improved human-likeness scores and positioned the model to behave more like a consultative HR assistant.

The final stage of the pipeline, **Low-Rank Adaptation (LoRA)**, was critical in transforming the model from a general assistant into a **domain-expert HR advisor**. Using a small but high-quality dataset focused on Malaysian employment regulations, workplace procedures, benefits administration, and statutory compliance, LoRA injected targeted knowledge into the model without requiring full parameter retraining. This parameter-efficient method allowed the model to respond accurately to region-specific HR scenarios such as calculating leave entitlements under Malaysian public holiday rules or citing clauses from the Employment Act 1955. The adapter-based tuning approach also makes the model future-proof, allowing for easy updates as policies evolve.

Quantitatively, the results confirmed the effectiveness of this layered approach. The model's **BLEU score improved from a baseline of 5.02% to 79.12%**, while **ROUGE-L increased from 35.91% to 92.23%** after all refinement stages. These dramatic improvements indicate a high degree of content overlap with reference answers and demonstrate the model's ability to replicate high-quality, human-authored HR responses. The **average response length also increased** appropriately, reflecting the model's ability to elaborate with sufficient detail without becoming verbose.

Qualitative evaluations further supported these gains. Human raters assessed model outputs across four key dimensions: fluency, relevance, helpfulness, and human-likeness, using a 5-point Likert scale. The final LoRA-enhanced model achieved **top-tier scores in fluency and relevance (5/5)**, and near-perfect marks in helpfulness and human-likeness. Reviewers consistently noted the model's ability to handle real-world HR scenarios with professionalism, clarity, and local accuracy, an outcome rarely achieved by general-purpose LLMs.

Importantly, these results were obtained using a **compact 123M-parameter model**, proving that **scaling is not the only path to capability**. By integrating modern Transformer architecture enhancements such as **RMSNorm**, **SwiGLU**, and **Rotary Position Embeddings (RoPE)** with focused training strategies, we demonstrated that even small-scale models can deliver high-impact, domain-specific functionality when optimized appropriately.

In summary, this study validates that **MiniMind-Dense functions as an effective, efficient, and trustworthy HR chatbot**, capable of providing Malaysian SMEs with around-the-clock assistance in areas such as policy explanation, benefits clarification, and procedural compliance. The modular nature of the training pipeline especially the use of LoRA adapters also means that this solution is **adaptable and maintainable over time**, ensuring long-term viability as legal frameworks and organizational needs evolve.

Ultimately, this work contributes a scalable framework for building **localized, purpose-specific language models** and offers a compelling alternative to massive generic LLMs by demonstrating that **precision, alignment, and efficiency** can be achieved through **intelligent training design rather than brute-force scaling**.

5.2. Reflection on Refinement Techniques and Architecture

This study illustrates the power of **combining architectural enhancements with a modular, multi-phase refinement**

strategy to create a domain-specialized language model that balances performance with efficiency. The development of **MiniMind-Dense** serves as a compelling case study for how thoughtful design rather than brute-force scaling can enable small models to operate effectively in complex, knowledge-intensive domains such as human resource management.

From an architectural perspective, the integration of **modern Transformer improvements** such as **RMSNorm**, **SwiGLU**, and **Grouped-Query Attention (GQA)** played a pivotal role in establishing a high-capacity, lightweight foundation. RMSNorm helped stabilize training dynamics, especially in smaller models where layer-wise variance can cause instability. SwiGLU, a gated activation function, introduced greater non-linear expressiveness without significantly increasing parameter count or computational cost. Meanwhile, GQA optimized the attention mechanism by reducing redundancy across multiple heads, allowing the model to process longer sequences more efficiently. Together, these enhancements enabled MiniMind-Dense to handle **extended HR queries** with greater contextual awareness and fewer resources, an essential trait for real-time applications in SME environments.

Beyond architecture, the staged refinement pipeline proved equally critical. Each phase addressed a different aspect of model alignment and capability:

- **Supervised Fine-Tuning (SFT)** served as the foundational corrective layer. By training on well-structured conversational data, the model quickly learned how to produce more complete, coherent, and task-aware responses. The quantitative metrics bear this out, BLEU nearly tripled after this stage alone, showing that SFT effectively reduced the domain gap between generic pretraining and applied HR use cases.
- **Reinforcement Learning from Human Feedback (RLHF)** using **Proximal Policy Optimization (PPO)** added an essential layer of human-aligned behavior. This phase helped the model optimize beyond factual correctness toward response qualities like **empathy, clarity, and professionalism**. Even though the parameter updates were relatively minor, the impact on response tone and perceived helpfulness was significant, confirming that even a light application of preference-based learning can dramatically enhance user experience.
- **Low-Rank Adaptation (LoRA)** emerged as one of the most transformative yet efficient techniques in the entire pipeline. By injecting a small number of task-specific parameters into selected layers, LoRA enabled **rapid specialization without overwriting the model's general capabilities**. The near-perfect overlap with reference answers, reflected in BLEU and ROUGE-L scores exceeding 79% and 92%, respectively proves that even a tiny module can guide the model toward **deep domain expertise**. In the context of HR, this means new laws, updated company policies, or industry-specific requirements can be incorporated simply by training a new LoRA adapter, without the need to retrain the entire base model.

This modularity is not just a technical convenience, it has profound practical implications. In real-world settings, especially among SMEs, resource limitations often prohibit frequent full-model updates. The ability to selectively refine parts of the model using LoRA or similar adapter methods makes it possible to maintain and evolve the system in a **cost-effective and agile** manner. Updates can be scheduled to reflect changes in government regulation or HR policies, allowing the AI assistant to remain legally compliant and practically useful over time.

Moreover, the success of this layered approach suggests a generalizable framework for future domain-specific SLM development. Rather than treating language models as monolithic systems that require retraining from scratch, our findings support a **compositional strategy**, one in which foundational linguistic capability is gradually shaped through targeted interventions, each one enhancing a different axis of performance.

In conclusion, the integration of modern architectural strategies with SFT, RLHF, and LoRA produced a model that is not only lightweight and efficient but also deeply aligned with its target domain. These findings reinforce the viability of a **layered, modular design philosophy** for domain adaptation particularly valuable for environments constrained by compute resources, data availability, or deployment scalability. MiniMind-Dense stands as a practical demonstration that **smart engineering choices and strategic refinement can substitute for sheer model scale**, and that high-impact, real-world applications of AI can be achieved without sacrificing accessibility or precision.

5.3. Implications for HR Domain Applications

The refined **MiniMind-Dense** model demonstrates strong potential for **real-world deployment in Malaysian HR environments**, particularly within small and medium-sized enterprises (SMEs) where staffing and legal expertise are often limited. Its ability to produce accurate, structured, and policy-aligned responses makes it an ideal candidate for roles such as an **internal HR chatbot**, **digital helpdesk assistant**, or even a **first-level query resolver** for employees navigating HR-related concerns. These use cases are particularly relevant for SMEs that may not have a dedicated HR department, enabling broader access to timely and compliant HR support **[33]**.

One of the most significant advantages of MiniMind-Dense lies in its **domain adaptation through localized training**. Because the model was explicitly refined on Malaysian HR dialogues, spanning legal compliance topics, workplace communication norms, and public-sector entitlements, it demonstrates a strong understanding of the regulatory and cultural landscape specific to Malaysia. As a result, the model requires **minimal customization or retraining** for deployment across local organizations, which significantly reduces onboarding time and cost. This **off-the-shelf usability** is especially attractive to SMEs, which often lack the technical expertise to fine-tune or maintain complex AI systems.

Furthermore, the model's **parameter-efficient architecture**, enabled by **Low-Rank Adaptation (LoRA)**, provides a scalable solution for continuous domain alignment. As labor laws and HR regulations evolve such as changes in leave entitlements, minimum wage, or employee rights, new LoRA adapters can be fine-tuned and integrated without the need to retrain the entire model. This enables **rapid updates and low-latency compliance integration**, which is critical for maintaining organizational accuracy and legal accountability in dynamic regulatory environments.

In addition to legal compliance, MiniMind-Dense can enhance **communication quality and consistency** across an organization. By standardizing the language, tone, and content of HR responses, it helps minimize misunderstandings, reduce bias, and ensure that all employees receive clear, equitable guidance. This is particularly valuable in multilingual and multicultural workplaces like those in Malaysia, where consistent messaging is essential for fairness and operational harmony.

Moreover, the model opens the door to **personalized employee engagement tools**. For example, it can be extended to assist in onboarding, policy explanation, performance review preparation, or benefits consultation automating repetitive tasks while ensuring alignment with organizational values and policies. Its **human-like tone**, achieved through RLHF, makes it approachable and suitable for interactions that require empathy and discretion, such as handling workplace grievances or personal leave inquiries.

Overall, the deployment of MiniMind-Dense offers tangible benefits in terms of **HR productivity, legal risk mitigation, employee satisfaction, and cost efficiency**. Its design lightweight, modular, and localized, aligns well with the operational realities of Malaysian SMEs, allowing them to **leverage the power of AI without the complexity or scale requirements typically associated with large language models**. By combining technical rigor with domain relevance, this solution contributes to a more inclusive and intelligent HR infrastructure in the national business landscape.

5.4. Future Work

While the current version of **MiniMind-Dense** demonstrates strong performance and domain alignment for Malaysian HR tasks, several promising directions could further enhance its capabilities, adaptability, and real-world impact. These potential extensions span both architectural innovations and functional improvements designed to expand the model's versatility, reduce hallucinations, and improve accessibility.

One major enhancement would be the integration of **Retrieval-Augmented Generation (RAG)**, a technique that supplements generative models with real-time document retrieval mechanisms **[5]**. By linking the model to a trusted document database such as company HR manuals, government policy archives, or up-to-date labor regulations, the assistant could retrieve and incorporate **verbatim excerpts or clauses** from relevant sources into its responses. This would substantially improve factual grounding, allowing the model to **quote authoritative sources directly**, rather than relying solely on internalized knowledge. RAG would not only reduce hallucinations but also increase user trust by improving traceability and auditability of model outputs in compliance-sensitive environments.

Another important direction is the development of **multilingual capabilities**, specifically to support both **Bahasa Malaysia (Malay)** and **English**, which are commonly used in Malaysian workplaces **[34]**. Currently, the model is optimized for English input and output, but many SMEs operate in bilingual contexts, where legal documents, employee contracts, and daily communication may alternate between languages. By implementing **translation-based fine-tuning** or conducting **bilingual training** on HR-related Malay corpora, the model could support **seamless code-switching or full-language toggling**, thus improving inclusivity and expanding its applicability across industries and employee demographics.

In terms of architecture, future versions of MiniMind-Dense could explore **scaling up the base model to larger parameter sizes**. For example, from 123M to 1B or 2B parameters to improve **reasoning depth, contextual fluency, and long-form coherence [35]**. Larger models generally exhibit better zero-shot generalization and can retain more complex policy structures or decision-making patterns. When combined with domain-specific fine-tuning and LoRA-style adapters, this scaling strategy could create highly capable assistants for more advanced HR applications, such as automated contract generation, internal policy auditing, or sentiment-aware employee feedback analysis.

Another valuable avenue for improvement is the implementation of **continual learning techniques** to enable the model to evolve over time without forgetting prior knowledge. Rather than retraining from scratch whenever new data becomes available, continual learning frameworks would allow the system to **incrementally update itself** based on new HR policies, legal amendments, or user feedback, while retaining stability and accuracy across previously learned content. This would support long-term maintainability and ensure that the assistant remains **relevant, legally up-to-date, and robust against concept drift**.

Lastly, broader **human-centered evaluation and deployment studies** could greatly enrich the model's real-world utility. Conducting **user experience (UX) research**, **longitudinal field testing**, and **ethnographic interviews** with HR personnel across different SME sectors would help refine the model's tone, response timing, and interaction style **[36]**. Feedback from diverse end users can also guide improvements in areas such as accessibility, bias mitigation, and the ethical handling of sensitive HR topics like complaints, terminations, or harassment cases.

In conclusion, by exploring these extensions: **retrieval grounding, multilingualism, scalability, continual learning, and UX-driven refinement**, MiniMind-Dense can evolve into an even more powerful and inclusive digital HR assistant. These future directions not only promise technical advancements but also reinforce the model's alignment with the practical needs of Malaysian businesses and the ethical standards of AI deployment in workplace settings.

Acknowledgements

The authors would like to thank colleagues from Tunku Abdul Rahman University of Management and Technology (TARUMT) who provided insights and expertise that greatly assisted the research, although they may not necessarily agree with all of the interpretations or conclusions of this paper. Special thanks to Prof. Lim Tong Ming for his supervision and guidance throughout the development of this work.

Author contributions

Darren Chai Xin Lun: Conceptualization, Methodology, Software, Field study, Data curation, Writing – Original draft preparation, Validation
Prof. Lim Tong Ming: Supervision, Methodological guidance, Writing – Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] J. Bersin, "The role of generative AI and large language models in HR," Industry blog article, Mar. 10, 2023

[2] J. Gong, MiniMind [Computer software], 2025.

[3] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: A method for automatic evaluation of machine translation," in Proc. 40th Annu. Meeting Assoc. Comput. Linguistics, 2002, pp. 311–318..

[4] *C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," in Proc. ACL-04 Workshop: Text Summarization Branches Out, 2004, pp. 74–81.*

[5] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," arXiv preprint arXiv:2005.11401, 2020.

[6] R. Gong, "Not leaving MSMEs behind in the AI race," Policy report,

Khazanah Research Institute, Oct. 30, 2024.

[7] D. Manickam, "Govt to develop localised large language model," News Article, Dec. 4, 2024.

[8] H. Zolkepli, A. Razak, K. Adha, and A. Nazhan, "MaLLaM - Malaysia Large Language Model," arXiv preprint arXiv:2401.14680v2, Jan. 2024

[9] . Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," Bioinformatics, vol. 36, no. 4, pp. 1234–1240, 2020.

[10] I. Beltagy, K. Lo, and A. Cohan, "SciBERT: A pretrained language model for scientific text," in Proc. EMNLP, 2019, pp. 3615–3620

[11] J. Su, Y. Lu, S. Pan, A. Murtadha, B. Wen, and Y. Liu, RoFormer: Enhanced transformer with rotary position embedding, arXiv:2104.09864v5, Nov. 8, 2023. [Online]. Available: https://arxiv.org/abs/2104.09864

[12] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M. A. Lachaux, T. Lacroix et al., "LLaMA: Open and efficient foundation language models," arXiv preprint arXiv:2302.13971, 2023.

[13] N. Shazeer, GLU variants improve transformer, arXiv:2002.05202, Feb. 12, 2020. [Online]. Available: https://arxiv.org/abs/2002.05202

[14] Z. Zhang et al., ReLU2 wins: Discovering efficient activation functions for sparse LLMs, arXiv:2402.03804v1, Feb. 6, 2024. [Online]. Available: https://arxiv.org/abs/2402.03804

[15] J. Ainslie, J. Lee-Thorp, M. de Jong, Y. Zemlyanskiy, F. Lebron, and S. Sanghai, "GQA: Training generalized multi-query transformer models from multi-head checkpoints," in Proc. 2023 Conf. Empirical Methods in Natural Language Processing (EMNLP), H. Bouamor, J. Pino, and K. Bali, Eds., Singapore, pp. 4895–4901, Dec. 2023. [Online]. Available: https://doi.org/10.18653/v1/2023.emnlp-main.298

[16] C. Fang, C. Qin, Q. Zhang, and K. Yao, "RecruitPro: A pretrained language model with skill-aware prompt learning for intelligent recruitment," in Proc. 29th ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD '23), Aug. 2023. [Online]. Available: https://doi.org/10.1145/3580305.3599894

[17] N. Otani, N. Bhutani, and E. Hruschka, Natural language processing for human resources: A survey, arXiv:2410.16498v2, Mar. 25, 2025. [Online]. Available: https://arxiv.org/abs/2410.16498v2

[18] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin et al., "Training language models to follow instructions with human feedback," arXiv preprint arXiv:2203.02155, 2022.

[19] J. Chen, X. Han, Y. Ma, X. Zhou, and L. Xiang, Unlock the correlation between supervised fine-tuning and reinforcement learning in training code large language models, arXiv:2406.10305v2, Dec. 17, 2024. [Online]. Available: https://arxiv.org/abs/2406.10305v2

[20] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang et al., "LoRA: Low-rank adaptation of large language models," arXiv preprint arXiv:2106.09685, 2022.

[21] R. Rafailov, A. Sharma, E. Mitchell, S. Ermon, C. D. Manning, and C. Finn, Direct preference optimization: Your language model is secretly a reward model, arXiv:2305.18290v3, Jul. 29, 2024. [Online]. Available: https://arxiv.org/abs/2305.18290v3

[22] M. Madanchian, "From recruitment to retention: AI tools for human resource decision-making," Appl. Sci., vol. 14, no. 24, p. 11750, 2024. [Online]. Available: https://doi.org/10.3390/app142411750

[23] A. Lacroux and C. Martin Lacroux, "Should I trust the artificial intelligence to recruit? Recruiters' perceptions and behavior when faced with algorithm-based recommendation systems during resume screening," Front. Psychol., vol. 13, p. 895997, 2022. [Online]. Available: https://doi.org/10.3389/fpsyg.2022.895997

[24] D. Zielinski, "How HR is using generative AI in performance management," Society for Human Resource Management, Aug. 8, 2023. [Online]. Available: https://www.shrm.org/topics-tools/news/technology/how-hr-using-generative-ai-performance-management

[25] S. K. Ho, "Amendments to Malaysia's Personal Data Protection Act 2010 (PDPA)," Deloitte Southeast Asia, Oct. 29, 2024. [Online]. Available: https://www.deloitte.com/southeast-asia/en/services/consulting-risk/perspectives/my-pdpa-amendments.html

[26] C. S. Seah, A. N. A. Nuar, Y. X. Loh, F. W. Jalaludin, H. Y. Foo, and L. L. Har, "Exploring the adoption of artificial intelligence in SMEs: An investigation into the Malaysian business landscape," Pacific Corporate Sustainability, vol. 2, no. 3, Article 35, 2023. [Online]. Available: https://doi.org/10.55092/pcs2023020035

[27] V. B. Parthasarathy, A. Zafar, A. Khan, and A. Shahid, "The ultimate guide to fine-tuning LLMs from basics to breakthroughs: An exhaustive review," arXiv preprint arXiv:2408.13296, 2024.

[28] S. Raschka, Build a Large Language Model (from scratch). Manning Publications, 2024.

[29] A. T. Leejoy, "Training language models to follow instructions with human feedback: A comprehensive review," Medium blog article, Mar. 2, 2025

[30] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347, 2017.

[31] *P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei, "Deep reinforcement learning from human preferences," in Adv. Neural Inf. Process. Syst., vol. 30, 2017*

[32] S. Chatterjee, N. P. Rana, K. Tamilmani, and A. Sharma, "The adoption of artificial intelligence in human resource management: Towards a research agenda," Int. J. Inf. Manage., vol. 52, p. 102019, 2020.

[33] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán et al., "Unsupervised cross-lingual representation learning at scale," arXiv preprint arXiv:1911.02116, 2020.

[34] S. Zhang, S. Roller, N. Goyal, M. Artetxe, M. Chen, S. Chen et al., "OPT: Open pre-trained transformer language models," arXiv preprint arXiv:2205.01068, 2022.

[35] M. Delange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis et al., "A continual learning survey: Defying forgetting in classification tasks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 7, pp. 3366–3385, 2021

[36] B. Heo, S. Park, D. Han, and S. Yun, Rotary position embedding for vision transformer, arXiv:2403.13298v2, Jul. 16, 2024. [Online]. Available: https://arxiv.org/abs/2403.13298