EXPLOITING ANONYMOUS ANNOTATIONS FOR CLUSTER-BASED LEARNING TO DEFER

Anonymous authors
Paper under double-blind review

ABSTRACT

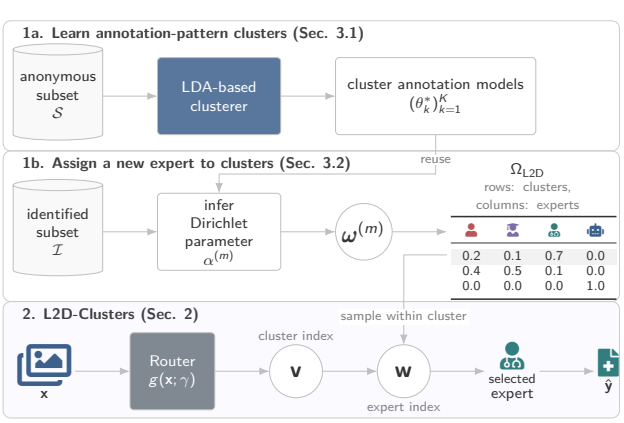
Learning to defer (L2D) is a prominent paradigm in human-AI cooperation because it leverages both the efficiency of machine learning models and the reliability of human experts. However, existing L2D methods are typically expert-specific, meaning that the same set of human experts annotating during training must also be present at deployment. Such a requirement limits the flexibility of L2D, particularly in dynamic real-world environments where expert availability can fluctuate due to leave, retirement, or the onboarding of new members. To accommodate such changes without redefining the router for every expert pool, we propose L2D-Clusters, a cluster-based L2D method that defers a query to a cluster, from which an expert (either a human or classifier) is selected to make the prediction. To construct these clusters while reducing the need for annotations linked to individual identities, we introduce a Bayesian mixed-membership model that learns reusable labelling components from anonymous annotations. The learnt model subsequently infers the cluster membership of a previously unseen expert using only a small identified annotation set without ground-truth labels. Experiments show that the learnt clusters group experts whose annotations agree on the same inputs and that L2D-Clusters outperforms expert-specific and population-based baselines in the small identified data regime, while expert-specific methods regain their advantage when abundant identified data are available.

1 INTRODUCTION

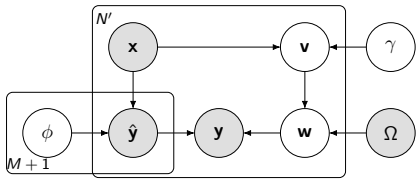
Advances in machine learning has led to several solutions to automate complex workflows to improve efficiency. While these solutions demonstrate exceptional efficiency when processing large amounts of information, they exhibit limitations, such as biases. In contrast, human experts, while less efficient at processing large volumes of data than machines, show remarkable adaptability to make reliable predictions. An ideal system should therefore combine machine efficiency with human reliability. *Learning to defer* (L2D) (Madras et al., 2018) is a promising approach to integrating human experts into the decision making pipeline to balance efficiency and reliability. By analysing annotated data, L2D estimates the reliability of human and machine decision-makers and routes each sample to balance predictive accuracy and operational cost.

Despite strong empirical performance, existing L2D methods remain largely expert-specific (Mozannar & Sontag, 2020; Verma & Nalisnick, 2022; Mao et al., 2023; Nguyen et al., 2025), requiring a fixed set of experts across training and deployment. Such a requirement poses a significant practical challenge because of real-world fluctuations in personnel due to the departure or onboarding of new team members. Since human experts often agree substantially on annotated samples (Nowak & Rüger, 2010), modelling each expert independently may introduce unnecessary redundancy and require more expert-specific data than needed. These observations motivate a different formulation of L2D: instead of learning to defer to specific individuals, can we learn to defer to groups of experts whose annotations agree on the same inputs? Realising this idea requires addressing two challenges. First, the conventional expert-specific L2D formulation must be extended to support deferral at the level of expert groups. Second, such groups must be discovered from annotation data in a manner that remains robust to changes in expert availability. This challenge further raises a fundamental question: can anonymous annotations be exploited to identify groups of experts with similar annotations while reducing the dependence on annotations linked to individual expert identities?

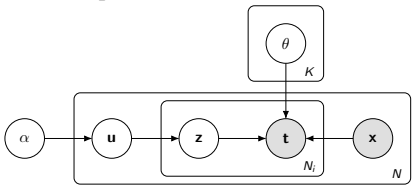
Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



(a) Diagram of the proposed two-stage clustering-based L2D.



(b) Graphical model of L2D-Clusters.



(c) Graphical model of the LDA-based clustering.

Figure 1: (a) The proposed clustering-based L2D consists of two stages: the first stage trains a clustering model on anonymous annotation data (1a) and assign experts to clusters (1b), and the second stage performs L2D on the clusters of human experts; (b) and (c) show the graphical models of L2D-Clusters and the LDA-based clustering with shaded nodes representing latent variables.

In this paper, we address both challenges. Specifically, (i) we propose a cluster-based L2D framework replacing expert-indexed deferral by cluster-indexed deferral to support deployment under expert onboarding and workforce turnover; (ii) we develop a Bayesian method for discovering groups of basic labelling functions from anonymous annotations and "softly" assigning each unseen expert to those groups from a set of limited identified annotations without requiring ground-truth labels; and (iii) we provide theoretical and empirical analyses demonstrating that the proposed framework improves sample-efficient adaptation to new experts and achieves superior performance in realistic onboarding scenarios.

2 L2D WITH CLUSTERS OF HUMAN EXPERTS

We consider an L2D setting consisting of an annotated dataset: $\mathcal{D} = \{(\mathbf{x}_n, (\hat{\mathbf{y}}_n^{(m)})_{m=1}^M, \mathbf{y}_n)\}_{n=1}^{N'}$, where $\mathbf{x}_n \in \mathcal{X}$ is an input sample, $\hat{\mathbf{y}}_n^{(m)} \in \mathcal{Y} = \{1, \ldots, C\}$ is the annotation made by a human expert indexed by $m \in \{1, \ldots, M\}$ and $\mathbf{y}_n \in \mathcal{Y}$ is the ground truth label. These human experts have been assigned to one of the K clusters with their probability of cluster assignment defined in the matrix $\Omega = \begin{bmatrix}\omega^{(1)} & \omega^{(2)} & \ldots & \omega^{(M)}\end{bmatrix}$, where $\omega^{(m)} \in \Delta^{K-1} = \{\omega^{(m)} | \omega^{(m)} \in [0,1]^K \wedge \mathbf{1}^\top \omega^{(m)} = 1\}$ is the probability vector representing the cluster assignment for expert m. Here, Ω is assumed to be given. Its estimation is presented in Section 3.2. The aim is to learn an L2D system that defers to one cluster and have one expert (either human or classifier) in that cluster make the final decision.

The desired system is analogous to a two-level hierarchical MoE (Jordan & Jacobs, 1994), in which the first level corresponds to expert clusters, while the second level consists of the human experts in each of those clusters. To include the classifier in an additional and separate cluster, the given cluster assignment probability matrix Ω is expanded from K to $K+1$ clusters as follows:

$$\Omega_{\text{L2D}} = \begin{bmatrix}\omega^{(1)} & \omega^{(2)} & \ldots & \omega^{(M)} & \mathbf{0} \\ 0 & 0 & \ldots & 0 & 1\end{bmatrix} \in \mathbb{R}^{(K+1)\times(M+1)}, \tag{1}$$

where the last column denotes the probability of cluster assignment for the classifier. We follow a generative approach to model the data generation process for $K+1$ clusters of experts (human and classifier) to learn an L2D system. Specifically, a decision can be generated as follows:

1. draw a sample: $\mathbf{x} \sim \Pr(\mathbf{x})$,
2. draw a cluster index: $\mathbf{v} \sim \Pr(\mathbf{v}|\mathbf{x};\gamma) = \text{Categorical}(\mathbf{v}|g(\mathbf{x};\gamma))$,
3. draw an expert's index from that cluster: $\mathbf{w} \sim \Pr(\mathbf{w}|\mathbf{v}) = \text{Categorical}\left(\mathbf{w}\,\middle|\,\Omega_{\text{L2D}}^\top \vec{\mathbf{e}}_v / \mathbf{1}^\top \Omega_{\text{L2D}}^\top \vec{\mathbf{e}}_v\right)$,
4. draw a "belief" of that expert: $\hat{\mathbf{y}}^{(w)} \sim \Pr(\hat{\mathbf{y}}|\mathbf{x},\mathbf{w};\phi) = \text{Categorical}(\hat{\mathbf{y}}|f(\mathbf{x};\phi^{(w)}))$,
5. draw a label from the expert's belief: $\mathbf{y} \sim \Pr(\mathbf{y}|\hat{\mathbf{y}},\mathbf{w}) = \text{Categorical}\left(\mathbf{y}|\hat{\mathbf{y}}^{(w)}\right)$,

where: $g(.;\gamma) : \mathcal{X} \to \Delta^K$ is a function, parameterised by γ, representing a router, $\vec{\mathbf{e}}_j$ is the j-th unit coordinate vector, and $f(.;\phi^{(m)}) : \mathcal{X} \to \Delta^{C-1}$ is a classification model representing an expert (including the classifier) indexed by $m \in \{1, \dots, M+1\}$. Note that the function representing each human expert, denoted as $f(.;\phi^{(m)}), \forall m \in \{1, \dots, M\}$, is simply a mapping table that returns an annotation $\hat{\mathbf{y}}^{(m)} = f(\mathbf{x};\phi^{(m)})$ given an input $\mathbf{x}$.

Fig. 1b shows the graphical model of this generative process. The objective is to learn the router's parameter by maximising the likelihood of the ground truth label given the system's predictions:

$$\begin{aligned}&\max_{\gamma,\phi^{(M+1)}} \sum_{n=1}^{N'} \ln \Pr(\mathbf{y}_n, \prod_{m=1}^{M+1} \hat{\mathbf{y}}_n^{(m)} | \mathbf{x}_n; \gamma, \Omega_{\text{L2D}}, \prod_{m=1}^{M+1} \phi^{(m)}) \\ &= \max_{\gamma,\phi^{(M+1)}} \sum_{n=1}^{N'} \ln \Pr(\mathbf{y}_n, | \mathbf{x}_n, \prod_{m=1}^{M+1} \hat{\mathbf{y}}_n^{(m)}; \gamma, \Omega_{\text{L2D}}) + \ln \Pr(\hat{\mathbf{y}}_n^{(M+1)} | \mathbf{x}_n; \phi^{(M+1)}),\end{aligned} \tag{2}$$

where the notation $\prod$ is abused to denote the joint probability of multiple random variables.

The first term of Eq. (2) is to train the router parameterised by γ, while the second term is to train the classifier. Thus, the annotation $\hat{\mathbf{y}}^{(M+1)}$ of the classifier must be modified accordingly. In the first term, it is the prediction of the classifier: $\hat{\mathbf{y}}^{(M+1)} = f(\mathbf{x};\phi^{(M+1)})$, so that the router can assess the performance of each expert (including humans and the classifier) to defer. In the second term, it is the ground truth label: $\hat{\mathbf{y}} = \mathbf{y}$ to train the classifier. To simplify the analysis, we omit the second term by training the classifier on ground truth labels separately, and consider the trained classifier as another un-learnable expert. Hence, Eq. (2) is simplified to consist of the first term only.

Directly optimising Eq. (2) is difficult due to the presence of latent variables $\mathbf{v}$ and $\mathbf{w}$. Hence, we employ the EM algorithm (Dempster et al., 1977) to optimise a lower-bound as follows:

$$Q_n = \mathbb{E}_{\Pr(\mathbf{v}_n, \mathbf{w}_n | \mathbf{x}_n, \mathbf{y}_n, \prod_{m=1}^{M+1} \hat{\mathbf{y}}_n^{(m)}, \gamma^{(\kappa)}, \Omega)} \left[\ln \Pr(\mathbf{y}_n, \prod_{m=1}^{M+1} \hat{\mathbf{y}}_n^{(m)}, \mathbf{v}_n, \mathbf{w}_n \middle| \mathbf{x}_n, \gamma, \Omega) \right]. \tag{3}$$

The E step is to calculate the posterior of the latent variables $\mathbf{v}$ and $\mathbf{w}$ (the one under the expectation in Eq. (3)), while the M step is to maximise Q_n in Eq. (3) w.r.t. γ. Appendix A.1 provides more details on the parameter inference of L2D-Clusters.

Constrain workload assignment Unconstrained L2D-Clusters may route most queries to high-performing expert clusters, yielding an imbalanced and cost-ineffective allocation. We therefore impose workload constraints, following *Probabilistic-L2D* (Nguyen et al., 2025), to regulate deferral. Specifically, the posterior of $\mathbf{v}$ and $\mathbf{w}$ is projected onto a space satisfying the workload constraint:

$$\begin{aligned}\{q^*(\mathbf{v}_n, \mathbf{w}_n)\}_{n=1}^{N'} &= \arg\min_q \tfrac{1}{N'} \sum_{n=1}^{N'} \text{KL}\left[q(\mathbf{v}_n, \mathbf{w}_n) \| p(\mathbf{v}_n, \mathbf{w}_n)\right] \\ &\text{subject to: } \boldsymbol{\varepsilon}_l \preceq \tfrac{1}{N'} \sum_{n=1}^{N'} \sum_{\mathbf{v}_n=1}^{K+1} q(\mathbf{v}_n, \mathbf{w}_n) \preceq \boldsymbol{\varepsilon}_u,\end{aligned} \tag{4}$$

where: $q(\mathbf{v}_n, \mathbf{w}_n)$ is the posterior satisfying the constraint, $p(\mathbf{v}_n, \mathbf{w}_n) = \Pr(\mathbf{v}_n, \mathbf{w}_n \middle| \mathbf{x}_n, \mathbf{y}_n, \prod_{m=1}^{M+1} \mathbf{t}_n^{(m)}, \gamma^{(\tau)}, \Omega)$ is the unconstrained posterior, $\boldsymbol{\varepsilon}_l, \boldsymbol{\varepsilon}_u \in [0,1]^{M+1}$ are hyper-parameters controlling the workload constraint, and $\preceq$ is the element-wise operator for less than or equal. Readers are referred to Appendix A.2 for further details.

Since the performance of L2D-Clusters depends on the cluster assignment matrix Ω, we therefore analyse the robustness of L2D-Clusters in Proposition 1 with its proof presented in Appendix B.

Proposition 1. *Let Ω^* be any reference cluster-assignment matrix, $\widehat{\Omega}$ be an estimated one, $s_{\min} \leq \min(\sum_m \Omega^*_{km}, \sum_m \hat{\Omega}_{km}) \forall k \in \{1, \dots, K\}$ denotes a lower bound on active cluster mass, L be the upperbound of the non-negative loss incurred by a prediction and ground truth, and $\mathcal{R}(g, \Omega)$ be the risk of L2D-Clusters for a router g and cluster assignment Ω. For any learnt router $\widehat{g}$, the following holds:*

$$\mathcal{R}(\widehat{g}, \widehat{\Omega}) - \mathcal{R}(g^*_{\Omega^*}, \Omega^*) \leq \tfrac{2L}{s_{\min}} \sum_{k=1}^{K} \|\widehat{\Omega}_k - \Omega^*_k\|_1 + \varepsilon_{\text{router}}, \tag{5}$$

where: $g^*_\Omega = \text{argmin}_g \mathcal{R}(g, \Omega)$ is the router that minimises the risk for a given Ω, and $\varepsilon_{\text{router}} \geq \mathcal{R}(\widehat{g}, \widehat{\Omega}) - \mathcal{R}(g^*_{\widehat{\Omega}}, \widehat{\Omega})$ is the generalisation error of the learnt router $\widehat{g}$.

Proposition 1 decomposes the excess risk of L2D-Clusters into two terms. The first term quantifies the cost of imperfect cluster assignment, while the second term, $\varepsilon_{\text{router}}$, captures the statistical and optimisation error of learning the router. Subsequent analysis on the first term will be presented in Section 3.2 to obtain an end-to-end guarantee for L2D-Clusters.

3 CLUSTERING HUMAN EXPERTS

One essential factor in L2D-Clusters, presented in Section 2, is the cluster assignment probability matrix Ω. Naively clustering the annotations $(\mathbf{t}^{(m)})_{m=1}^{M}|\mathbf{x}$ is incorrect because doing so clusters individual annotations, not the annotators who produced them. We instead cluster annotators, not annotations, using a generative model, inspired by Latent Dirichlet Allocation (LDA) (Blei et al., 2003) from topic modelling. Each annotator plays the role of a document, and each annotation plays the role of a word in it. As a document is modelled as a mixture of topics, we model each annotator as a mixture over basic "annotation styles", with their annotations generated from that mixture. And because LDA can discover topics from a collection of documents without knowing who authored each one, our proposed model can also discover annotation-style clusters from a set of annotations without knowing which annotator produced each of them.

Throughout the paper, "annotation style" denotes a labelling function, which is a component of the mixture. Cluster membership is therefore determined by which components best explain an expert's observed annotations via annotation agreement rather than any external notion of annotation style.

3.1 MODEL ANONYMOUS ANNOTATIONS WITH LATENT DIRICHLET ALLOCATION

The anonymous-annotated dataset is denoted as: $\mathcal{S} = \cup_{i=1}^{D}\mathcal{S}_i$ with $\mathcal{S}_i = \{(\mathbf{x}_{ij}, \mathbf{t}_{ij})\}_{j=1}^{N_i}$, where $\mathbf{t}_{ij} \in \mathcal{Y}$ is an anonymous annotation associated with the input sample $\mathbf{x}_{ij} \in \mathcal{X}$. Each subset $\mathcal{S}_i$ represents a collection of annotated samples collected randomly, which can either be from the same or different annotators. Each annotated sample in $\mathcal{S}_i$ can be generated as follows:

1. draw a mixture of basic annotation styles from a Dirichlet prior: $\mathbf{u}_{ij} \sim \mathrm{Dir}(\mathbf{u}|\alpha)$,
2. draw a sample from data distribution: $\mathbf{x}_{ij} \sim \Pr(\mathbf{x})$,
3. draw a basic annotation style: $\mathbf{z}_{ij} \sim \Pr(\mathbf{z}|\mathbf{u}_{ij}) = \mathrm{Categorical}(\mathbf{z}|\mathbf{u}_{ij})$,
4. draw an annotation: $\mathbf{t}_{ij} \sim \Pr(\mathbf{t}|\mathbf{x}_{ij}, \mathbf{z}_{ij}; (\theta_k)_{k=1}^{K}) = \mathrm{Categorical}(\mathbf{t}|h(\mathbf{x}_{ij}; \theta_{\mathbf{z}_{ij}}))$,

where $\alpha \in \mathbb{R}_{+}^{K}$ represents the concentration parameter of the Dirichlet prior, $h(.;\theta_k) : \mathcal{X} \to \Delta^{C-1}$ is a model parameterised by θ_k to represent a basic annotation style (i.e., a classifier). The graphical model of this generative process is shown in Fig. 1c.

The objective is to learn the set of model parameters $(\theta_k)_{k=1}^{K}$, each representing a basic annotation style, by maximising the likelihood on the observed anonymous annotation dataset as follows:

$$(\theta_k^*)_{k=1}^{K} = \mathrm{argmax}_{(\theta_k)_{k=1}^{K}} \sum_{i=1}^{D}\sum_{j=1}^{N_i} \ln \Pr(\mathbf{t}_{ij}|\mathbf{x}_{ij}, \alpha, (\theta_k)_{k=1}^{K}). \tag{6}$$

The optimisation in Eq. (6) is similar to the variational EM algorithm in the conventional LDA, except that multinomial topics are replaced by classification models (see Appendix E).

Annotation diversity To learn the clustering model in Eq. (6), the anonymous set $\mathcal{S}$ must contain enough representative instances where the K underlying labelling functions $h(\cdot;\theta_k)$ disagree. If two clusters' true annotation functions are near-identical on $\mathcal{S}$ (i.e., $\mathbb{E}_{\mathcal{S}_i}\mathbb{E}_{\mathbf{x}\sim\mathcal{S}_i}[\mathrm{KL}(h(\mathbf{x};\theta_k)\|h(\mathbf{x};\theta_{k'}))]$ is small), then no clusterer, including a Bayes-optimal one, can separate them from finite samples. This leads to cluster collapsing. To reduce the risk, we set the Dirichlet prior α to a large value (e.g., $\alpha = \mathbf{1}\alpha_0 \wedge \alpha_0 > 1$). While this does not increase diversity, it encourages the LDA-based model to exploit available annotation variability more evenly across clusters.

3.2 ASSIGN TEST-TIME HUMAN EXPERTS INTO CLUSTERS

At test time, the learnt LDA-based model (i.e., $(\theta_k^*)_{k=1}^{K}$) is used to process a small *identified* set of data annotated by that human expert, denoted as $\mathcal{I}_m = \{(\mathbf{x}_j^{(m)}, \mathbf{t}_j^{(m)})\}_{j=1}^{N_m}$. The aim is to infer the mixture of annotation styles $\mathbf{u}$ that explains the annotations in $\mathcal{I}_m$. Casting $\mathcal{I}_m$ as an N_m-word document and each annotation style as a topic, this is exactly the problem of inferring the topic mixture of a test document in LDA. It is therefore carried out by the same variational inference procedure used during training (see Appendix E.2):

$$\rho^{(m)}, \zeta_{1:N_m}^{(m)} = \arg\min_{\rho,\zeta_{1:N_m}} \mathrm{KL}[q(\mathbf{u},\mathbf{z}|\rho,\zeta_{1:N_m})\|\Pr(\mathbf{u},\mathbf{z}|\mathcal{I}_m,\alpha,(\theta_k^*)_{k=1}^{K})]. \tag{7}$$

The cluster membership probability can then be interpreted as the mean of $\mathrm{Dir}(\mathbf{u}|\rho^{(m)})$ as follows:

$$\hat{\omega}^{(m)} = \mathbb{E}_{q(\mathbf{u}|\rho^{(m)})}[\mathbf{u}] = \rho^{(m)}/\mathbf{1}^{\top}\rho^{(m)}. \tag{8}$$

The probability vector of cluster assignment $\hat{\omega}_k^{(m)}$ in Eq. (8) of a test-time human expert m can be stacked column-wise to form a matrix $\hat{\Omega} = \begin{bmatrix}\hat{\omega}^{(1)} & \hat{\omega}^{(2)} & \dots & \hat{\omega}^{(M)}\end{bmatrix} \in \mathbb{R}^{K\times M}$ that represents the human clusters presented in Section 2. When performing downstream tasks, such as deferring to the cluster indexed by k (see Section 2), one can simply sample a random human expert with a probability vector proportional to the k-th row of the cluster matrix $\hat{\Omega}$. The whole procedure is visually illustrated in Fig. 1a and shown in Algorithm 1 (see Appendix M).

We analyse the deviation between the exact posterior $\omega^{(m)} = \mathbb{E}_{\Pr(\mathbf{u}|\mathcal{I}_m,\alpha,(\theta_k^*)_{k=1}^K)}[\mathbf{u}]$ and the estimate $\hat{\omega}^{(m)}$ in Proposition 2 (see the proof in Appendix F.1).

Proposition 2. *The cluster membership inferred from $\rho^{(m)}$ satisfies*

$$\|\hat{\omega}^{(m)} - \omega^{(m)}\|_1 \leq \min\left(2, \sqrt{2\,\mathrm{KL}[q(\mathbf{u},\mathbf{z}|\rho^{(m)},\zeta_{1:N_m}^{(m)})\|\Pr(\mathbf{u},\mathbf{z}|\mathcal{I}_m,\alpha,(\theta_k^*)_{k=1}^K)]}\right). \quad (9)$$

Proposition 2 shows that the error induced is bounded by the KL divergence. This KL divergence is defined in Eq. (7) representing the variational gap for expert m, and zero when the VI approximation in Eq. (7) is exact. In other words, the error depends on the quality of the optimisation that infers $\rho^{(m)}$. When the KL divergence is greater than 2, the bound is simplified to the universal simplex bound $\|\hat{\omega}^{(m)} - \omega^{(m)}\|_1 \leq 2$.

The result in Proposition 2 can also be integrated into Proposition 1 to yield an end-to-end bound on the excess risk of L2D-Clusters. In Proposition 1, we define the reference Ω^* as the exact posterior: $\Omega^* = \begin{bmatrix}\omega^{(1)} & \omega^{(2)} & \dots & \omega^{(M)}\end{bmatrix}$. Then, $\sum_{k=1}^K \|\widehat{\Omega}_k - \Omega_k^*\|_1 = \sum_{m=1}^M \|\hat{\omega}^{(m)} - \omega^{\mathrm{post}(m)}\|_1$, which Proposition 2 bounds.

Corollary 1. *Let $\mathrm{KL}_{\max} = \max_{m\in\{1,\dots,M\}} \mathrm{KL}[q(\mathbf{u},\mathbf{z}|\rho^{(m)},\zeta_{1:N_m}^{(m)})\|\Pr(\mathbf{u},\mathbf{z}|\mathcal{I}_m,\alpha,(\theta_k^*)_{k=1}^K)]$ be the worst-case variational gap over the M experts. The excess risk of L2D-Clusters is bounded as:*

$$\mathcal{R}(\widehat{g},\widehat{\Omega}) - \mathcal{R}(g_{\Omega^*}^*,\Omega^*) \leq \tfrac{2LM}{s_{\min}} \times \min\left(2, \sqrt{2\,\mathrm{KL}_{\max}}\right) + \varepsilon_{\mathrm{router}}.$$

Corollary 1 (proof in Appendix F.2) contains two terms, each representing the optimisation quality in each stage. The bound is linearly proportional to the number of experts M when the variational gap is non-zero. When the LDA-based clustering is exact, the bound reduces to the error induced by the learnt router only.

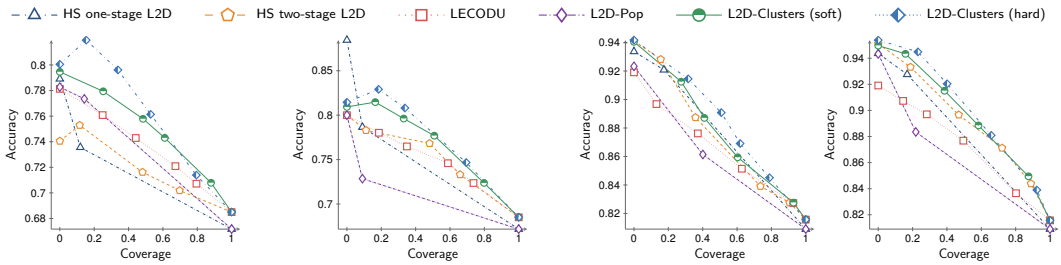
4 Experiments

We introduce a practical benchmark in Section 4.1 for L2D evaluation. We compare the performance of L2D-Clusters with existing L2D methods in Section 4.2. We also evaluate the performance of our LDA-based expert clustering method in Section 4.3.

4.1 Practical L2D benchmark

Existing benchmarks in L2D mostly follow the human-specific setting, in which each expert provides a substantial number of annotations. While these benchmarks are widely used, they fail to reflect practical scenarios such as expert onboarding, where human experts typically contribute only a limited number of samples. Although recent works have explored settings with incomplete or "missing" annotations (Hemmer et al., 2023; Nguyen et al., 2025), their training procedures assume a sufficiently large volume of data per expert and typically involve only a small number of experts (usually less than 5). In contrast, real-world datasets often comprise many experts, each contributing only a few annotations (e.g., 10 in CIFAR-10N (Wei et al., 2022) and 200 in ImageNet-1.2k (Steyvers et al., 2022)). Furthermore, most synthetic benchmarks in L2D simulate annotation errors using class-dependent instance-independent label noise, biasing the evaluation towards the approaches that directly incorporate this assumption into their modelling (e.g., IFD (Strong et al., 2026)). Moreover, such an overly-simplistic modelling of annotation mistakes does not align with empirical observations, as real-world label noise is frequently both class- and instance-dependent (Carneiro, 2024).

We, therefore, propose a new benchmark that reflects real-world annotation dynamics. Specifically, synthetic experts are simulated to make annotation errors under a class- and instance-dependent



(a) CIFAR-100/100 img. (b) CIFAR-100/200 img. (c) Chaoyang/50 img. (d) Chaoyang/100 img.

Figure 2: Comparison of coverage - accuracy curves between different standard L2D methods (i.e., single-stage and two-stage), L2D-Pop, LECODU and L2D-Clusters on a variety of datasets, and different number of identified/context samples N'.

Table 1: The means and standard deviations of area under the coverage - accuracy curve ($\times 100$) of many L2D methods with different numbers of context samples.

	CIFAR-100 (6 experts)		**dopanim**		**Chaoyang**	
Context size N'	100	200	100	200	50	100
One-stage	71.07 ± 2.34	73.87 ± 1.97	69.42 ± 1.04	69.42 ± 1.05	87.57 ± 0.96	87.97 ± 0.94
Two-stage	71.80 ± 1.55	75.02 ± 1.53	73.35 ± 1.00	74.03 ± 1.00	87.44 ± 0.85	89.31 ± 0.84
LECODU	73.60 ± 1.98	74.81 ± 1.72	74.00 ± 1.00	73.67 ± 0.99	86.46 ± 0.85	85.57 ± 0.85
IFD	29.70 ± 1.12	31.19 ± 1.02	35.97 ± 0.09	36.54 ± 0.09	43.38 ± 0.69	44.76 ± 0.70
L2D-Pop	73.05 ± 2.65	70.59 ± 2.17	65.23 ± 1.52	65.08 ± 1.52	85.81 ± 1.73	86.09 ± 1.70
L2D-Clusters (soft)	75.07 ± 1.06	76.68 ± 1.05	74.61 ± 0.72	75.09 ± 0.72	87.81 ± 0.85	89.69 ± 0.83
L2D-Clusters (hard)	$\mathbf{76.17} \pm 1.50$	$\mathbf{77.31} \pm 1.48$	$\mathbf{78.12} \pm 0.98$	$\mathbf{78.20} \pm 0.97$	$\mathbf{88.51} \pm 0.87$	$\mathbf{90.09} \pm 0.85$

label noise (Xia et al., 2020). Each expert contributes fewer than 200 training annotations, emulating realistic onboarding conditions. Additional evaluations are also performed in departure settings, in which a subset of experts available during training is no longer present at test time.

4.2 Learning to Defer with Small Context Sets

We employ the clustering method presented in Section 3.2 to obtain the cluster assignment matrix Ω to train an L2D-Clusters system. We use the benchmark in Section 4.1 to evaluate. We analyse coverage–accuracy curves, where coverage denotes the fraction of test samples handled by the classifier and accuracy denotes the final prediction performance.

We report two variants of our method: "soft" clustering probability, and its "hard" version (i.e., each human is assigned to only a single cluster by rounding the probability vector to a one-hot vector). Here, we use $K = 3$ when training our clustering model as it shows the optimal perplexity on the hold-out validation samples. In addition, as the proposed L2D-Clusters is stochastic in terms of selecting a human expert in a cluster to make decisions, we, therefore, calculate the average prediction accuracy weighted by the assignment probability of each human expert belonging to the selected cluster: $\mathsf{accuracy}_{\mathrm{L2D-Clusters}}(\mathbf{z} = k) = \sum_{m=1}^{M+1} \Omega'_{km} \mathbb{I}_{(\mathbf{t}^{(m)}=\mathbf{y})}$, where: Ω'_{km} is the normalised probability of expert m in the cluster k can be selected to make decisions (see Eqs. (1) and (8)) and $\mathbb{I}$ is the indicator function. This ensures a fair comparison.

Experiment setting We simulate three pairs of synthetic experts on CIFAR-100, each introducing 5%, 20%, and 35% annotation errors under class- and instance-dependent label noise. These experts are denoted as 5A, 5B, 20C, 20D, 35E, and 35F, where the numbers denote their labelling errors, while the letters indicate different random seeds. Our method trains the LDA-based clustering model on the full training set with anonymous annotations. During evaluation, the LDA-based model assigns cluster probability for each expert based on the *identified* annotations made by the corresponding expert (without ground truth). A similar simulation is also applied on the dataset dopanim. For the real-world dataset Chaoyang, we use the 3 human pathologists provided.

Baselines We consider two types of baselines: *(i)* the conventional human-specific L2D, including one-stage (Mozannar et al., 2023; Verma et al., 2023) and two-stage L2D methods (Madras

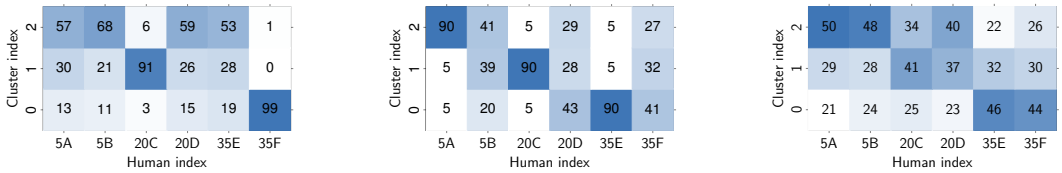
Table 2: (a) Performance of different L2D methods on the same 100-samples context or identified set in the setting of quitting or retirement: (left column) 4 experts by removing one 5-percent-noise-rate expert and one 20-percent-noise-rate expert, and (right column) 4 experts by removing both the 5-percent-noise-rate experts; (b) the adjusted Rand index (ARI) of expert clustering.

(a) Area under the coverage-accuracy curve for retirement

Method	**5B-20D-35E-35F**	**20C-20D-35E-35F**
One-stage	68.34 ± 2.73	70.41 ± 2.65
Two-stage	69.77 ± 1.98	71.25 ± 1.83
LECODU	72.01 ± 1.97	73.78 ± 1.73
IFD	33.08 ± 1.03	59.76 ± 1.24
L2D-Pop	70.68 ± 2.27	71.64 ± 2.39
L2D-Cluster (soft)	75.02 ± 1.10	**74.17** ± 1.11
L2D-Cluster (hard)	**75.58** ± 1.45	74.13 ± 1.57

(b) Unsupervised clustering performance

	Adjusted Rand Index			
Num. clusters	**2**	**3**	**4**	**5**
L2D-Pop	0	-0.06	0.29	-0.11
IFD	0.24	0.44	0.29	-0.11
LDA-based	0.24	0.44	0.44	1



(a) L2D-Pop - $K = 3$ (b) IFD - $K = 3$ (c) Ours - $K = 3$

Figure 3: Clustering results on the simulation of 6 synthetic experts on CIFAR-100: each column represents the probability of cluster assignment obtained from a context set of $N' = 100$ samples annotated by a human expert. Each human expert is simulated to annotate with 5%, 20% and 35% of mistakes at different random seeds (i.e., A, B, C, D, E, F), following the instance-dependent label noise. Ideally, each pair will be grouped into the same cluster. Note that the heat-map colour is relatively scaled to enhance the visualisation.

et al., 2018; Mao et al., 2023; Nguyen et al., 2025), and *(ii)* expert-agnostic L2D, including LECODU (Zhang et al., 2024), L2D-Pop (Tailor et al., 2024). When there are enough training samples: (i.e., N' is sufficiently large), the human-specific L2D methods will outperform L2D-Clusters. Our interest is more in the low-data regime (e.g., small *context* sets with ground-truth labels). For a fair comparison, we train LECODU directly on the anonymous dataset and set the deferral to only one random expert. The other baseline, L2D-Pop, however, can only defer to one specific human at a time. To adapt to the setting of multiple available human experts at test time, for each sample, we calculate the deferral probabilities of all available human experts, and consider the expert with the maximum deferral probability. In addition, the one-stage L2D (Verma & Nalisnick, 2022) and L2D-Pop (Tailor et al., 2024) do not provide a mechanism to control the coverage. Hence, we can only report a single point beside the two extreme points of coverages of zero and one. We do not provide qualitative results for IFD (Strong et al., 2026) because IFD relies on class-dependent instance-independent annotation behaviour, and hence, has poor performance in our instance-dependent benchmark.

Results Fig. 2 shows the *coverage-accuracy* curves of the four baselines and L2D-Clusters on various datasets, each with different numbers of annotator-identified samples N'. In general, both the human-specific baselines (one-stage and two-stage) and L2D-Pop overfit to small training *context* sets, resulting in low performance. Since LECODU defers to randomly selected experts, its coverage–accuracy curve typically aligns with the average performance obtained across deferral rates. In contrast, the proposed L2D-Clusters (both soft and hard) outperforms these baselines, demonstrating its capability of dealing with real-world annotated datasets when only a small amount of training samples are associated with the (pseudo) identities of human experts. Although our proposed clustering method presented in Section 3.1 can handle small N' in its inference, further reducing N' makes the training of L2D methods difficult due to the insufficient number of training samples. For example, using fewer than 100 samples may leave some classes in CIFAR-100 unrepresented. For larger N', the performance of L2D-Clusters also improves, but is lower than the human-specific one-stage and two-stage L2D methods (see Appendix C). This is expected because at larger N', the traditional one-stage and two-stage L2D baselines have enough information about each specific human expert, and hence, perform better than the L2D-Clusters.

We also report the quantitative results as the area under the coverage-accuracy curve across all methods. The results in Table 1 show that our proposed L2D-Clusters outperforms all the baselines, in which the hard version of L2D-Clusters has the highest performance across all the practical settings when the size of *identified*/*context* set, N', is small.

Large number of human experts We simulate a total of 100 synthetic experts on CIFAR-100, each with an error rate varying from 5% to 50%. The results shown in Fig. 5 are consistent with our previous findings, in which our proposed framework, including both LDA-based clustering and L2D-Clusters, outperforms the baselines and state-of-the-art methods in the onboarding setting (e.g., small context sets). Please refer to Appendix D for further details.

Departure and onboarding setting Table 2a shows the results in a departure setting where some experts are removed at testing. In general, L2D-Clusters variants consistently show strong performance compared to the baselines.

4.3 Clustering on anonymously-annotated data

Baselines We consider two baselines: L2D-Pop (Tailor et al., 2024) and IFD (Strong et al., 2026). These two methods obtain human representation either via learning, as in L2D-Pop, or hand-crafting, as in IFD. The representations of all human experts extracted from these two baselines are then clustered and compared with our LDA-based clustering method presented in Section 3.1. Specifically, the whole model of L2D-Pop is trained on the *context* (i.e., with ground-truth labels) data of all available human experts to extract the representations of those experts. For IFD, the statistics of each *context* set associated with each human expert are summarised via counting, and the mean and variance of the obtained Beta posterior are used to represent each human expert. The set of representations extracted from each baseline is then clustered by the Fuzzy C-Means algorithm (Dunn, 1973; Bezdek, 2013), which is a variant of the K-Means algorithm. Note that other clustering algorithms, such as Gaussian mixture models, may also be used. However, due to the high-dimensional nature of the representations obtained from these baselines (e.g., in the range of hundreds) coupled with a small number of samples (e.g., in our setting, the number of human experts is less than 10), learning full covariance matrices of a Gaussian mixture model is challenging.

Experiment setting For CIFAR-100, we follow the same setting of Section 4.2 to simulate three pairs of synthetic experts. For L2D-Pop and IFD, each uses *context* data to extract expert representations, which are then clustered using Fuzzy C-Means. Further details are in Appendix N.

Qualitative results The clustering heatmap in Fig. 3 shows that both the baselines, including L2D-Pop and IFD, do not consistently group each pair of synthetic human experts into the same cluster, while our proposed clustering method largely recovers the annotation similarity of each pair of synthetic experts. This outcome is expected: L2D-Pop exhibits poor performance in low-data regimes (here, $N' = 100$ samples) due to its reliance on data abundance, while IFD assumes class-dependent instance-independent annotation and hence, performs poorly under the realistic instance-dependent label noise. In contrast, our method (Fig. 3c), although not grouping expert pairs into perfectly separated clusters, reveals a smooth transition of the cluster assignment probabilities across pairs with a similar trend observed for $K \in \{2, 4, 5\}$ in Appendix G. This observation also reflects the annotation diversity discussed in Section 3.1, where annotation data with small disagreement ($\approx 30\%$) introduces ambiguity in forming distinct clusters. Nevertheless, within each cluster, paired experts exhibit nearly identical probabilities, demonstrating our model's ability to capture annotation similarity – a property absent in L2D-Pop and IFD.

Quantitative results We treat expert pairs (5A, 5B), (20C, 20D), and (35E, 35F) as reference clusters and use the *Adjusted Rand Index* (ARI) (Hubert & Arabie, 1985) for evaluation. As shown in Table 2b, our proposed LDA-based clustering method performs on par with or outperforms the two baselines, despite relying solely on the *identified* set of human annotations (without ground-truth), while the baselines need the *context* annotation set (with ground-truth). Also, at $K = 5$ clusters (Fig. 6 in Appendix G), our method correctly assigns all three pairs to distinct clusters while leaving two clusters empty, yielding a perfect ARI score. In contrast, the Fuzzy C-Means applied to the two baselines forces each expert into a separate cluster, yielding a negative ARI.

We also perform several ablation studies: the number of clusters in Appendix K, sensitivity analysis of the number of clusters in Appendix I and the Dirichlet prior's parameter α_0 in Appendix J, and additional results of clustering on real-world datasets in Appendix H.

Scope of LDA-based clustering As the LDA-based model observes only anonymous annotations, it groups experts by agreement and disagreement of annotations, rather than by the latent annotation patterns. We observe this with synthetic experts (5A, 20A) and (5B, 20B), where each pair shares an annotation style (A or B) but differs in error rate (5% vs 20%). The model recovers the groups (5A, 5B) and (20A, 20B), reflecting the high overlap of their annotations. This behaviour is, in fact, well aligned with the objective of L2D.

5 Related work

This section reviews L2D approaches. An extension of the related work, including clustering crowdsource workers and consensus from crowd is presented in Appendix O.

Learning to defer has recently attracted research interest due to its inherent capability to make autonomous decisions or selectively defer to a human expert when lacking sufficient confidence (Madras et al., 2018; Mozannar & Sontag, 2020; Keswani et al., 2021; Verma & Nalisnick, 2022; Verma et al., 2023; Mozannar et al., 2023; Nguyen et al., 2025). However, a prevalent characteristic among most existing L2D methodologies is their reliance on a human-specific deferral strategy, which often fails to account for the dynamic availability of human experts encountered in real-world deployment scenarios, distinct from development environments. To our best knowledge, only the data-driven L2D-Pop (Tailor et al., 2024) and hand-crafted IFD (Strong et al., 2026) circumvent this limitation by learning the representation of each human expert and training L2D models conditioned on such representation. Although these human-agnostic approaches increase the flexibility of L2D, both require a sufficient *context* set containing not only the annotations provided by that expert, but also the ground truth associated with each sample. Another human-agnostic L2D is LECODU (Zhang et al., 2024), in which a human expert is randomly selected. In contrast, our clustering model presented in Section 3.1 can be trained on anonymous annotation data which is more accessible in practice. Our proposed clustering approach does not need a *context* dataset (with ground truth) per human expert, but only an *identified* dataset (without ground truth) to assign that human expert to one of the clusters. Furthermore, the L2D-Clusters approach developed in Sec. 2 introduces a cluster-based deferral mechanism and chooses one expert within the group being selected for the final decision. Such deferral mechanism flexibility enables L2D-Clusters to readily adapt to diverse real-world scenarios (e.g., absence, retirement or onboarding), while simultaneously demonstrating performance comparable to human-specific L2D methods.

6 Discussion and conclusion

This paper introduces a cluster-based L2D method that defers queries to groups of experts, addressing the challenges posed by dynamic and rotating workforces. To construct expert clusters, we develop an LDA-based model that uncovers groups of experts with similar annotations from anonymous annotations, enabling efficient onboarding of new experts with limited data. The proposed approach is supported by theoretical analysis and extensive empirical evaluation in realistic settings. Empirically, the learnt clusters align well with known similarities in annotator performance, and when integrated with cluster-based L2D, consistently outperform expert-specific and expert-agnostic L2D baselines.

Our approach offers several practical advantages. By relying solely on anonymous annotations, it substantially reduces the data-collection burden associated with tracking individual expert identities and aligns naturally with privacy constraints in deployments. The cluster-based modelling is also well suited to dynamic expert pools, where individuals can leave or join the workforce.

The work has important limitations. As a data-centric approach, clustering quality depends on sufficient diversity in anonymous annotations. When experts exhibit highly similar behaviour, distinct clusters may become indistinguishable, leading to cluster merging or collapse. This is an identifiability limitation rather than a modelling deficiency, but it can constrain downstream deferral performance. A second characteristic is that the LDA-based model captures annotation co-occurrence, and hence, groups experts by the similarity of the annotations they produce. This is well aligned with the deferral objective, since an expert's value at test time depends on whether their annotation is correct. Future work includes analysing data to quantify annotation diversity, and modelling richer dependencies among clusters via more expressive models such as the correlated topic models (Blei & Lafferty, 2006).

AI USE STATEMENT

In this work, we used generative AI tools for:

- refining writing based on our own drafts,
- identifying and correcting grammatical errors,
- providing critical ingredients for proving mathematical claims,
- verifying the correctness of mathematical expressions and proofs, and
- providing feedback on research methodology or experiments.

We have not used generative AI tools for:

- developing theoretical models or conceptual frameworks,
- formulating mathematical claims,
- assisting in the writing of proofs,
- proposing or refining hypotheses, and
- implementing methods.

The AI-assisted tasks that are not applicable in our paper are:

- generate synthetic datasets,
- assist with translation,
- clean and reformat datasets,
- support qualitative and thematic data analysis,
- interpret results.

Additionally, the following AI-assisted tasks are neither used by us nor inapplicable to our work:

- Formulate questions for surveys or interviews,
- create or modify scientific figures or images,
- suggest experimental parameters,
- create or edit software code,
- creation of artifacts,
- draft parts of a research paper,
- transcribe recordings of research material,
- summarise or analyse existing literature,
- discover research topics or identify gaps,
- brainstorming,
- sourcing/searching for information,
- edit a research paper to improve readability,
- identify relevant literature,
- format references,
- suggest a structure for a research paper,
- propose a title or keywords for a research paper.

We have reviewed all AI-assisted work. Specifically, we manually verified the suggested writing revision to ensure its accuracy and appropriateness to our initial writing. We derived and wrote our proofs on physical papers before typing them into the electronic document. We also manually checked the feedback of our research methodology and experiments before revising to ensure they align with our research objectives and are scientifically sound. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

Ethics statement

Although our work is in the field of human-AI collaboration, it does not involve human subjects or any sensitive data. The annotations used in our experiments are either synthetically generated or publicly available from public datasets.

Reproducibility statement

The code to reproduce the results in this paper is attached in the Supplementary Materials. The hyper-parameters used and experimental setup for reproducibility are included in the Appendices.

References

Björn Barz and Joachim Denzler. Do we train on test data? Purging CIFAR of near-duplicates. *Journal of Imaging*, 6(6):41, 2020.

Valerio Basile. It's the end of the gold standard as we know it: Leveraging non-aggregated data for better evaluation and explanation of subjective tasks. In *International Conference of the Italian Association for Artificial Intelligence*, pp. 441–453. Springer, 2020.

James C Bezdek. *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.

David Blei and John Lafferty. Correlated topic models. In *Advances in Neural Information Processing Systems*, volume 18, pp. 147, 2006.

David M Blei and Michael I Jordan. Variational methods for the Dirichlet process. In *International Conference on Machine Learning*, pp. 12, 2004.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

Gustavo Carneiro. *Machine Learning with Noisy Labels: Definitions, Theory, Techniques and Solutions*. Elsevier, 2024.

Jonathan Chang, Sean Gerrish, Chong Wang, Jordan Boyd-Graber, and David Blei. Reading tea leaves: How humans interpret topic models. In *Advances in Neural Information Processing Systems*, volume 22, 2009.

Alexander Philip Dawid and Allan M Skene. Maximum likelihood estimation of observer error-rates using the EM algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28(1):20–28, 1979.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1): 1–22, 1977.

Naihao Deng, Xinliang Frederick Zhang, Siyang Liu, Winston Wu, Lu Wang, and Rada Mihalcea. You are what you annotate: Towards better models through annotator representations. In *ACL EMNLP*, 2023.

J. C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3):32–57, 1973. doi: 10.1080/01969727308546046.

Patrick Hemmer, Lukas Thede, Michael Vössing, Johannes Jakubik, and Niklas Kühl. Learning to defer with limited expert predictions. In *AAAI Conference on Artificial Intelligence*, volume 37, pp. 6002–6011, 2023.

Marek Herde, Denis Huseljic, Lukas Rauch, and Bernhard Sick. dopanim: A dataset of doppelganger animals with noisy annotations from multiple humans. In *Advances in Neural Information Processing Systems*, volume 37, pp. 51085–51117, 2024.

Matthew Hoffman, Francis Bach, and David Blei. Online learning for latent dirichlet allocation. In *Advances in Neural Information Processing Systems*, volume 23, 2010.

Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985.

Michael I Jordan and Robert A Jacobs. Hierarchical mixtures of experts and the em algorithm. *Neural computation*, 6(2):181–214, 1994.

Sanjay Kairam and Jeffrey Heer. Parting crowds: Characterizing divergent interpretations in crowd-sourced annotation tasks. In *ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pp. 1637–1648, 2016.

Markelle Kelly, Alex Boyd, Sam Showalter, Mark Steyvers, and Padhraic Smyth. Bayesian inference for correlated human experts and classifiers. In *International Conference on Machine Learning*, 2025.

Vijay Keswani, Matthew Lease, and Krishnaram Kenthapadi. Towards unbiased and accurate deferral to multiple experts. In *AAAI/ACM Conference on AI, Ethics, and Society*, pp. 154–165, 2021.

Soda Marem Lo and Valerio Basile. Hierarchical clustering of label-based annotator representations for mining perspectives. In *ECAI Workshop on Perspectivist Approaches to NLP*, 2023.

David Madras, Toni Pitassi, and Richard Zemel. Predict responsibly: improving fairness and accuracy by learning to defer. In *Advances in Neural Information Processing Systems*, 2018.

Anqi Mao, Christopher Mohri, Mehryar Mohri, and Yutao Zhong. Two-stage learning to defer with multiple experts. In *Advances in Neural Information Processing Systems*, volume 36, pp. 3578–3606, 2023.

Hussein Mozannar and David Sontag. Consistent estimators for learning to defer to an expert. In *International Conference on Machine Learning*, pp. 7076–7087. PMLR, 2020.

Hussein Mozannar, Hunter Lang, Dennis Wei, Prasanna Sattigeri, Subhro Das, and David Sontag. Who should predict? Exact algorithms for learning to defer to humans. In *International Conference on Artificial Intelligence and Statistics*, pp. 10520–10545. PMLR, 2023.

An T Nguyen, Matthew Lease, and Byron C Wallace. Explainable modeling of annotations in crowdsourcing. In *International Conference on Intelligent User Interfaces*, pp. 575–579, 2019.

Cuong Nguyen, Thanh-Toan Do, and Gustavo Carneiro. Probabilistic learning to defer: Handling missing expert's annotations and controlling workload distribution. In *International Conference on Learning Representations*, 2025.

Stefanie Nowak and Stefan Rüger. How reliable are annotations via crowdsourcing: A study about inter-annotator agreement for multi-label image annotation. In *International Conference on Multimedia Information Retrieval*, pp. 557–566, 2010.

Masa-Aki Sato. Online model selection based on the variational bayes. *Neural computation*, 13(7): 1649–1681, 2001.

Samuel Showalter, Alex J Boyd, Padhraic Smyth, and Mark Steyvers. Bayesian online learning for consensus prediction. In *International Conference on Artificial Intelligence and Statistics*, pp. 2539–2547. PMLR, 2024.

Mark Steyvers, Heliodoro Tejeda, Gavin Kerrigan, and Padhraic Smyth. Bayesian modeling of human–AI complementarity. *Proceedings of the National Academy of Sciences*, 119(11): e2111547119, 2022.

Joshua Strong, Pramit Saha, Yasin Ibrahim, Cheng Ouyang, and Alison Noble. Identity-free deferral for unseen experts. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=4YG9ufFg58.

Dharmesh Tailor, Aditya Patra, Rajeev Verma, Putra Manggala, and Eric Nalisnick. Learning to defer to a population: A meta-learning approach. In *International Conference on Artificial Intelligence and Statistics*, pp. 3475–3483. PMLR, 2024.

Panagiotis A Traganitis and Georgios B Giannakis. Identifying dependent annotators in crowdsourcing. In *Asilomar Conference on Signals, Systems, and Computers*, pp. 1276–1280. IEEE, 2022.

Rajeev Verma and Eric Nalisnick. Calibrated learning to defer with one-vs-all classifiers. In *International Conference on Machine Learning*, pp. 22184–22202. PMLR, 2022.

Rajeev Verma, Daniel Barrejón, and Eric Nalisnick. Learning to defer to multiple experts: Consistent surrogate losses, confidence calibration, and conformal ensembles. In *International Conference on Artificial Intelligence and Statistics*, pp. 11415–11434. PMLR, 2023.

Jiaheng Wei, Zhaowei Zhu, Hao Cheng, Tongliang Liu, Gang Niu, and Yang Liu. Learning with noisy labels revisited: A study using real-world human annotations. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=TBWA6PLJZQm.

Maximilian Wich, Hala Al Kuwatly, and Georg Groh. Investigating annotator bias with a graph-based approach. In *Workshop on Online Abuse and Harms*, pp. 191–199, 2020.

Xiaobo Xia, Tongliang Liu, Bo Han, Nannan Wang, Mingming Gong, Haifeng Liu, Gang Niu, Dacheng Tao, and Masashi Sugiyama. Part-dependent label noise: Towards instance-dependent label noise. In *Advances in Neural Information Processing Systems*, volume 33, pp. 7597–7610, 2020.

Zheng Zhang, Wenjie Ai, Kevin Wells, David Rosewarne, Thanh-Toan Do, and Gustavo Carneiro. Learning to complement and to defer to multiple users. In *European Conference on Computer Vision*, pp. 144–162. Springer, 2024.

Weizhong Zhao, James J Chen, Roger Perkins, Zhichao Liu, Weigong Ge, Yijun Ding, and Wen Zou. A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinformatics*, 16(13):S8, 2015.

Chuang Zhu, Wenkai Chen, Ting Peng, Ying Wang, and Mulan Jin. Hard sample aware noise robust learning for histopathology image classification. *IEEE Transactions on Medical Imaging*, 2022.

A Detailed derivation for L2D-Clusters

A.1 Parameter inference

The data modelling of the L2D-Clusters presented in Section 2 can also be illustrated via the graphical model shown in Fig. 1b.

The objective function of L2D-Clusters is to learn the gating model's parameter γ and the classifier's parameter ϕ_{M+1} by maximising data likelihood on observed annotated data:

$$\begin{aligned}&\max_{\gamma,\phi^{(M+1)}} \sum_{n=1}^{N'} \ln \Pr(\mathbf{y}_n, \prod_{m=1}^{M+1} \hat{\mathbf{y}}_n^{(m)} \Big| \mathbf{x}_n; \gamma, \Omega_{\text{L2D}}, \prod_{m=1}^{M+1} \phi^{(m)}) \\ &= \max_{\gamma,\phi^{(M+1)}} \sum_{n=1}^{N'} \ln \Pr(\mathbf{y}_n, | \mathbf{x}_n, \prod_{m=1}^{M+1} \hat{\mathbf{y}}_n^{(m)}; \gamma, \Omega_{\text{L2D}}) + \ln \Pr(\hat{\mathbf{y}}_n^{(M+1)} | \mathbf{x}_n; \phi^{(M+1)}). \quad (2)\end{aligned}$$

The first term in Eq. (2) learns the gating model $g(.;\gamma)$, while the second term learns the classifier $f(.;\phi_{M+1})$. We, therefore, train a classifier on ground truth data, and consider such a trained classifier as a fixed expert.

Due to the presence of latent variables: *(i)* the index of the cluster being selected $\mathbf{v}$, and *(ii)* the index of the expert (either human or classifier) being selected $\mathbf{w}$, evaluating the likelihood in Eq. (2) is difficult. We, therefore, employ the EM algorithm to optimise. In the EM algorithm, we first define completed-data log-likelihood as follows:

$$Q_n = \mathbb{E}_{\Pr(\mathbf{v}_n, \mathbf{w}_n | \mathbf{x}_n, \mathbf{y}_n, \prod_{m=1}^{M+1} \hat{\mathbf{y}}_n^{(m)}, \gamma^{(\tau)}, \Omega)} \left[\ln \Pr(\mathbf{y}_n, \prod_{m=1}^{M+1} \hat{\mathbf{y}}_n^{(m)}, \mathbf{v}_n, \mathbf{w}_n | \mathbf{x}_n, \gamma, \Omega) \right]. \quad (3)$$

E step calculates the posterior of latent variables:

$$\Pr(\mathbf{v}_n, \mathbf{w}_n | \mathbf{x}_n, \mathbf{y}_n, \prod_{m=1}^{M+1} \hat{\mathbf{y}}_n^{(m)}, \gamma^{(\tau)}, \Omega) \propto \Pr(\mathbf{y}_n | \prod_{m=1}^{M+1} \hat{\mathbf{y}}_n^{(m)}, \mathbf{w}_n) \Pr(\mathbf{w}_n | \mathbf{v}_n, \Omega) \Pr(\mathbf{v}_n | \mathbf{x}_n, \gamma^{(\tau)}).$$

M step maximises the completed-data log-likelihood w.r.t. γ as: $\gamma^{(\tau+1)} \leftarrow \text{argmax}_\gamma \frac{1}{N'} \sum_{i=1}^{N'} Q_n$.

Compared to Probabilistic-L2D (Nguyen et al., 2025), our formulation is simpler because the posteriors of both $\mathbf{v}$ and $\mathbf{w}$ can be obtained exactly in closed-form.

A.2 Workload constraint in L2D-Clusters

As mentioned in Section 4.2, instead of constraining on the clusters of experts, we constrain the workload assignment over all experts in all the clusters. The purpose is to have a fair comparison with the baselines.

The constrained optimisation to control workload across all experts (including both human and the classifier) can be written as follows:

$$\begin{aligned}&\{q^*(\mathbf{v}_n, \mathbf{w}_n)\}_{n=1}^{N'} = \underset{q}{\text{argmin}} \frac{1}{N'} \sum_{n=1}^{N'} \text{KL} \left[q(\mathbf{v}_n, \mathbf{w}_n) \| \Pr(\mathbf{v}_n, \mathbf{w}_n | \mathbf{x}_n, \mathbf{y}_n, \prod_{m=1}^{M+1} \mathbf{t}_n^{(m)}, \gamma^{(\tau)}, \Omega) \right] \\ &\text{s.t.: } \varepsilon_l \preceq \frac{1}{N'} \sum_{n=1}^{N'} \sum_{\mathbf{v}_n=1}^{K+1} q(\mathbf{v}_n, \mathbf{w}_n) \preceq \varepsilon_u. \end{aligned} \quad (4)$$

We apply the Lagrangian to solve the constrained optimisation in Eq. (4). The Lagrangian of the constrained optimisation can be written as follows:

$$
\begin{aligned}
\mathsf{L} &= \frac{1}{N'}\sum_{n=1}^{N'} \mathrm{KL}\left[q(\mathbf{v}_n,\mathbf{w}_n)\|\Pr(\mathbf{v}_n,\mathbf{w}_n|\mathbf{x}_n,\mathbf{y}_n,\prod_{m=1}^{M+1}\mathbf{t}_n^{(m)},\gamma^{(\tau)},\Omega)\right] \\
&\quad + \boldsymbol{\lambda}_u^\top\left[\frac{1}{N'}\sum_{n=1}^{N'}\sum_{\mathbf{v}_n=1}^{K+1} q(\mathbf{v}_n,\mathbf{w}_n) - \boldsymbol{\varepsilon}_u\right] + \boldsymbol{\lambda}_l^\top\left[\boldsymbol{\varepsilon}_l - \frac{1}{N'}\sum_{n=1}^{N'}\sum_{\mathbf{v}_n=1}^{K+1} q(\mathbf{v}_n,\mathbf{w}_n)\right] \\
&\quad + \lambda_q\left[\sum_{\mathbf{v}_n=1}^{K+1}\sum_{\mathbf{w}_n=1}^{M+1} q(\mathbf{v}_n,\mathbf{w}_n) - 1\right] \\
&= \frac{1}{N'}\sum_{n=1}^{N'}\sum_{\mathbf{z_n}=1}^{K+1}\sum_{\mathbf{w}_n=1}^{M+1} q(\mathbf{v}_n,\mathbf{w}_n)\left[\ln q(\mathbf{v}_n,\mathbf{w}_n) - \ln\Pr(\mathbf{v}_n,\mathbf{w}_n|\mathbf{x}_n,\mathbf{y}_n,\prod_{m=1}^{M+1}\mathbf{t}_n^{(m)},\gamma^{(\tau)},\Omega)\right] \\
&\quad + \boldsymbol{\lambda}_u^\top\left[\frac{1}{N'}\sum_{n=1}^{N'}\sum_{\mathbf{v}_n=1}^{K+1} q(\mathbf{v}_n,\mathbf{w}_n) - \boldsymbol{\varepsilon}_u\right] + \boldsymbol{\lambda}_l^\top\left[\boldsymbol{\varepsilon}_l - \frac{1}{N'}\sum_{n=1}^{N'}\sum_{\mathbf{v}_n=1}^{K+1} q(\mathbf{v}_n,\mathbf{w}_n)\right] \\
&\quad + \lambda_q\left[\sum_{\mathbf{v}_n=1}^{K+1}\sum_{\mathbf{w}_n=1}^{M+1} q(\mathbf{v}_n,\mathbf{w}_n) - 1\right],
\end{aligned}
\tag{10}
$$

where: $\boldsymbol{\lambda}_u$ and $\boldsymbol{\lambda}_l \in \mathbb{R}_+^{M+1}$ are the Lagrange multipliers.

Taking the derivative w.r.t. each element in the matrix $q(\mathbf{v}_n = k, \mathbf{w}_n = m)$ with $k \in \{1,\ldots,K\}$ and $m \in \{1,\ldots,M+1\}$ gives:

$$
\begin{aligned}
&\frac{\partial \mathsf{L}}{\partial q(\mathbf{v}_n = k, \mathbf{w}_n = m)} \\
&= \frac{1}{N'}\Bigg[\ln q(\mathbf{v}_n = k, \mathbf{w}_n = m) - \ln\Pr(\mathbf{v}_n = k, \mathbf{w}_n = m|\mathbf{x}_n,\mathbf{y}_n,\prod_{m=1}^{M+1}\mathbf{t}_n^{(m)},\gamma^{(\tau)},\Omega) \\
&\quad +1 + \boldsymbol{\lambda}_{um} - \boldsymbol{\lambda}_{lm}] + \lambda_q.
\end{aligned}
\tag{11}
$$

Applying the KKT condition gives:

$$
\begin{aligned}
&\frac{\partial \mathsf{L}}{\partial q(\mathbf{v}_n = k, \mathbf{w}_n = m)} = 0 \\
&\Leftrightarrow \ln q(\mathbf{v}_n = k, \mathbf{w}_n = m) - \ln\Pr(\mathbf{v}_n = k, \mathbf{w}_n = m|\mathbf{x}_n,\mathbf{y}_n,\prod_{m=1}^{M+1}\mathbf{t}_n^{(m)},\gamma^{(\tau)},\Omega) \\
&\qquad + 1 + \boldsymbol{\lambda}_{um} - \boldsymbol{\lambda}_{lm} + N'\lambda_q = 0 \\
&\Leftrightarrow \ln q(\mathbf{v}_n = k, \mathbf{w}_n = m) = \ln\Pr(\mathbf{v}_n = k, \mathbf{w}_n = m|\mathbf{x}_n,\mathbf{y}_n,\prod_{m=1}^{M+1}\mathbf{t}_n^{(m)},\gamma^{(\tau)},\Omega) \\
&\qquad - 1 - \boldsymbol{\lambda}_{um} + \boldsymbol{\lambda}_{lm} + \text{const.}.
\end{aligned}
\tag{12}
$$

Or:

$$
q(\mathbf{v}_n = k, \mathbf{w}_n = m) = \frac{1}{Z_n(\boldsymbol{\lambda}_u,\boldsymbol{\lambda}_l)}\frac{\Pr(\mathbf{v}_n = k, \mathbf{w}_n = m|\mathbf{x}_n,\mathbf{y}_n,\prod_{m=1}^{M+1}\mathbf{t}_n^{(m)},\gamma^{(\tau)},\Omega)}{\exp(\boldsymbol{\lambda}_{um} - \boldsymbol{\lambda}_{lm} + 1)}, \tag{13}
$$

where: $Z_n(\boldsymbol{\lambda}_u,\boldsymbol{\lambda}_l)$ is the normalisation constant defined as follows:

$$
Z_n(\boldsymbol{\lambda}_u,\boldsymbol{\lambda}_l) = \sum_{k=1}^{K+1}\sum_{m=1}^{M+1}\frac{\Pr(\mathbf{v}_n = k, \mathbf{w}_n = m|\mathbf{x}_n,\mathbf{y}_n,\prod_{m=1}^{M+1}\mathbf{t}_n^{(m)},\gamma^{(\tau)},\Omega)}{\exp(\boldsymbol{\lambda}_{um} - \boldsymbol{\lambda}_{lm} + 1)}. \tag{14}
$$

Substituting $q(\mathbf{v}_n = k, \mathbf{w}_n = m)$ obtained in Eq. (13) into the Lagrangian in Eq. (10) gives:

$$\begin{aligned}\mathsf{L} = &\frac{1}{N'}\sum_{n=1}^{N'}\sum_{k=1}^{K+1}\sum_{m=1}^{M+1} q(\mathbf{v}_n = k, \mathbf{w}_n = m)\left[-\ln Z_n(\boldsymbol{\lambda}_u, \boldsymbol{\lambda}_l) - \boldsymbol{\lambda}_{um} + \boldsymbol{\lambda}_{lm} - 1\right] \\ &+ \frac{1}{N'}\sum_{n=1}^{N'}\sum_{k=1}^{K+1}\sum_{m=1}^{M+1}(\boldsymbol{\lambda}_{um} - \boldsymbol{\lambda}_{lm})q(\mathbf{v}_n = k, \mathbf{w}_n = m) \\ &+ \sum_{m=1}^{M+1}\boldsymbol{\lambda}_{lm}\boldsymbol{\varepsilon}_{lm} - \boldsymbol{\lambda}_{um}\boldsymbol{\varepsilon}_{um} + \text{const.}\end{aligned} \tag{15}$$

Or:

$$\begin{aligned}\mathsf{L} &= -\boldsymbol{\lambda}_u^\top\boldsymbol{\varepsilon}_u + \boldsymbol{\lambda}_l^\top\boldsymbol{\varepsilon}_l - 1 - \frac{1}{N'}\sum_{n=1}^{N'}\ln Z_n(\boldsymbol{\lambda}_u, \boldsymbol{\lambda}_l) \\ &= -\boldsymbol{\lambda}_u^\top\boldsymbol{\varepsilon}_u + \boldsymbol{\lambda}_l^\top\boldsymbol{\varepsilon}_l - 1 - \frac{1}{N'}\sum_{n=1}^{N'}\ln\sum_{k=1}^{K+1}\sum_{m=1}^{M+1}\frac{\Pr(\mathbf{v}_n = k, \mathbf{w}_n = m|\mathbf{x}_n, \mathbf{y}_n, \prod_{m=1}^{M+1}\mathbf{t}_n^{(m)}, \gamma^{(\tau)}, \Omega)}{\exp(\boldsymbol{\lambda}_{um} - \boldsymbol{\lambda}_{lm} + 1)}.\end{aligned} \tag{16}$$

According to the duality, the Lagrange multipliers $\boldsymbol{\lambda}_u$ and $\boldsymbol{\lambda}_l$ can be obtained by maximising the Lagrangian, which is equivalent to the following:

$$\begin{aligned}\boldsymbol{\lambda}_u^*, \boldsymbol{\lambda}_l^* = &\operatorname*{argmin}_{\boldsymbol{\lambda}_u, \boldsymbol{\lambda}_l \geq 0}\boldsymbol{\lambda}_u^\top\boldsymbol{\varepsilon}_u - \boldsymbol{\lambda}_l^\top\boldsymbol{\varepsilon}_l + 1 \\ &+ \frac{1}{N'}\sum_{n=1}^{N'}\ln\sum_{k=1}^{K+1}\sum_{m=1}^{M+1}\frac{\Pr(\mathbf{v}_n = k, \mathbf{w}_n = m|\mathbf{x}_n, \mathbf{y}_n, \prod_{m=1}^{M+1}\mathbf{t}_n^{(m)}, \gamma^{(\tau)}, \Omega)}{\exp(\boldsymbol{\lambda}_{um} - \boldsymbol{\lambda}_{lm} + 1)}.\end{aligned} \tag{17}$$

The optimisation can then be obtained by applying an off-the-shelf optimisation solver, such as non-negative projection.

B Robust analysis for L2D-Clusters under cluster-assignment perturbation

In this section, we present the details of the proof for Proposition 1. For convenience, we restate the proposition here.

Proposition 1. *Let Ω^* be any reference cluster-assignment matrix, $\widehat{\Omega}$ be an estimated one, $s_{\min} \leq \min(\sum_m \Omega^*_{km}, \sum_m \hat{\Omega}_{km}) \forall k \in \{1, \dots, K\}$ denotes a lower bound on active cluster mass, L be the upperbound of the non-negative loss incurred by a prediction and ground truth, and $\mathcal{R}(g, \Omega)$ be the risk of L2D-Clusters for a router g and cluster assignment Ω. For any learnt router $\widehat{g}$, the following holds:*

$$\mathcal{R}(\widehat{g}, \widehat{\Omega}) - \mathcal{R}(g^*_{\Omega^*}, \Omega^*) \leq \tfrac{2L}{s_{\min}} \textstyle\sum_{k=1}^{K} \|\widehat{\Omega}_k - \Omega^*_k\|_1 + \varepsilon_{\text{router}}, \tag{5}$$

where: $g^*_\Omega = \operatorname{argmin}_g \mathcal{R}(g, \Omega)$ is the router that minimises the risk for a given Ω, and $\varepsilon_{\text{router}} \geq \mathcal{R}(\widehat{g}, \widehat{\Omega}) - \mathcal{R}(g^*_{\widehat{\Omega}}, \widehat{\Omega})$ is the generalisation error of the learnt router $\widehat{g}$.

Proof. Let $\Omega \in \mathbb{R}_+^{K \times M}$ denote the unnormalised cluster-assignment matrix used by L2D-Clusters in Section 2. For a cluster indexed by k, the probability that a human expert indexed by m is selected can be written as $\frac{\Omega_{km}}{\sum_{j=1}^{M} \Omega_{kj}}, \forall m \in \{1, \dots, M\}$. Note that this formula holds because the classifier (or the expert indexed by $(M+1)$) is in a separate cluster indexed by $(K+1)$. Its unnormalised probability in one of any K clusters is zero, and hence, when normalising as above, it can be omitted.

Let $g : \mathcal{X} \to \Delta^K$ be a router, where $g_k(\mathbf{x}; \gamma)$ is the probability of deferring the sample $\mathbf{x}$ to cluster k. Let $\ell_m(x, y) \in [0, L]$ denote a bounded loss incurred when expert m predicts on the annotated sample $(\mathbf{x}, \mathbf{y})$. The L2D risk induced by g and Ω is

$$\mathcal{R}(g, \Omega) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}}\Big[\underbrace{\sum_{k=1}^{K} g_k(\mathbf{x}; \gamma) \sum_{m=1}^{M} \frac{\Omega_{km}}{\sum_{j=1}^{M} \Omega_{kj}} \ell_m(\mathbf{x}, \mathbf{y})}_{K \text{ clusters of } M \text{ humans}} + \underbrace{g_{K+1}(\mathbf{x}; \gamma) \ell_{M+1}(\mathbf{x}, \mathbf{y})}_{\text{separate cluster of classifier alone}}\Big]. \tag{18}$$

Let Ω^* be a reference cluster-assignment matrix and $\widehat{\Omega}$ be the inferred one. No property of Ω^* is used below other than the cluster-mass condition that follows, so the result holds for any admissible reference. Assume that every active cluster has non-vanishing mass, i.e.,

$$\sum_{m=1}^{M} \Omega^*_{km} \geq s_{\min}, \qquad \sum_{m=1}^{M} \widehat{\Omega}_{km} \geq s_{\min}, \qquad \forall k \in \{1, \dots, K\}, \tag{19}$$

for some $s_{\min} > 0$. Then, the risk difference evaluated on the same router model g with two different clustering assignment matrices $\widehat{\Omega}$ and Ω^* can be written as follows:

$$\begin{aligned}
&\mathcal{R}(g, \widehat{\Omega}) - \mathcal{R}(g, \Omega^*) \\
&= \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}}\left[\sum_{k=1}^{K} g_k(\mathbf{x}; \gamma) \sum_{m=1}^{M} \left(\frac{\widehat{\Omega}_{km}}{\sum_{j=1}^{M} \widehat{\Omega}_{kj}} - \frac{\Omega^*_{km}}{\sum_{j=1}^{M} \Omega^*_{kj}}\right) \ell_m(\mathbf{x}, \mathbf{y})\right] \\
&= \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}}\left[\sum_{k=1}^{K} g_k(\mathbf{x}; \gamma) \sum_{m=1}^{M} \left(\frac{\widehat{\Omega}_{km} - \Omega^*_{km}}{\sum_{j=1}^{M} \widehat{\Omega}_{kj}} - \Omega^*_{km}\left(\frac{1}{\sum_{j=1}^{M} \Omega^*_{kj}} - \frac{1}{\sum_{j=1}^{M} \widehat{\Omega}_{kj}}\right)\right) \ell_m(\mathbf{x}, \mathbf{y})\right].
\end{aligned} \tag{20}$$

Applying the inequality $a - b \leq |a| + |b|$ gives:

$$
\begin{aligned}
&\mathcal{R}(g, \widehat{\Omega}) - \mathcal{R}(g, \Omega^*) \\
&\leq \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}} \left[\sum_{k=1}^{K} g_k(\mathbf{x};\gamma) \sum_{m=1}^{M} \left(|\frac{\widehat{\Omega}_{km} - \Omega^*_{km}}{\sum_{j=1}^{M} \widehat{\Omega}_{kj}}| + \Omega^*_{km} |\frac{1}{\sum_{j=1}^{M} \Omega^*_{kj}} - \frac{1}{\sum_{j=1}^{M} \widehat{\Omega}_{kj}}| \right) \ell_m(\mathbf{x}, \mathbf{y}) \right] \\
&\leq L\mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}} \left[\sum_{k=1}^{K} g_k(\mathbf{x};\gamma) \sum_{m=1}^{M} \left(\frac{|\widehat{\Omega}_{km} - \Omega^*_{km}|}{s_{\min}} + \frac{\Omega^*_{km}}{\sum_{j=1}^{M} \Omega^*_{kj}} \frac{|\sum_{j=1}^{M} \widehat{\Omega}_{kj} - \Omega^*_{kj}|}{\sum_{j=1}^{M} \widehat{\Omega}_{kj}} \right) \right] \\
&\qquad \text{(since } 0 \leq \ell_m(\mathbf{x}, \mathbf{y}) \leq L\text{)} \\
&\leq L\mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}} \left[\sum_{k=1}^{K} g_k(\mathbf{x};\gamma) \left(\frac{\sum_{m=1}^{M} |\widehat{\Omega}_{km} - \Omega^*_{km}|}{s_{\min}} + \frac{|\sum_{j=1}^{M} \widehat{\Omega}_{kj} - \Omega^*_{kj}|}{\sum_{j=1}^{M} \widehat{\Omega}_{kj}} \right) \right] \\
&\qquad = L\mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}} \left[\sum_{k=1}^{K} g_k(\mathbf{x};\gamma) \left(\frac{1}{s_{\min}} \|\widehat{\Omega}_k - \Omega^*_k\|_1 + \frac{|\sum_{j=1}^{M} \widehat{\Omega}_{kj} - \Omega^*_{kj}|}{\sum_{j=1}^{M} \widehat{\Omega}_{kj}} \right) \right] \\
&\leq L\mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}} \left[\sum_{k=1}^{K} g_k(\mathbf{x};\gamma) \left(\frac{1}{s_{\min}} \|\widehat{\Omega}_k - \Omega^*_k\|_1 + \frac{\sum_{j=1}^{M} |\widehat{\Omega}_{kj} - \Omega^*_{kj}|}{s_{\min}} \right) \right] \\
&\leq \frac{2L}{s_{\min}} \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}} \left[\sum_{k=1}^{K} g_k(\mathbf{x};\gamma) \|\widehat{\Omega}_k - \Omega^*_k\|_1 \right] \\
&\leq \frac{2L}{s_{\min}} \sum_{k=1}^{K} \|\widehat{\Omega}_k - \Omega^*_k\|_1 \qquad \text{(note that: } g_k(\mathbf{x};\gamma) \leq 1\text{)}.
\end{aligned}
\tag{21}
$$

Furthermore, let $g^*_{\Omega^*} \in \operatorname{argmin}_{g\in\mathcal{G}} \mathcal{R}(g, \Omega^*)$ and $g^*_{\widehat{\Omega}} \in \operatorname{argmin}_{g\in\mathcal{G}} \mathcal{R}(g, \widehat{\Omega})$, where $\mathcal{G}$ denotes the class of feasible routers, including any workload constraints. If the learned router $\widehat{g}$ satisfies

$$
\mathcal{R}(\widehat{g}, \widehat{\Omega}) - \mathcal{R}(g^*_{\widehat{\Omega}}, \widehat{\Omega}) \leq \varepsilon_{\text{router}}, \tag{22}
$$

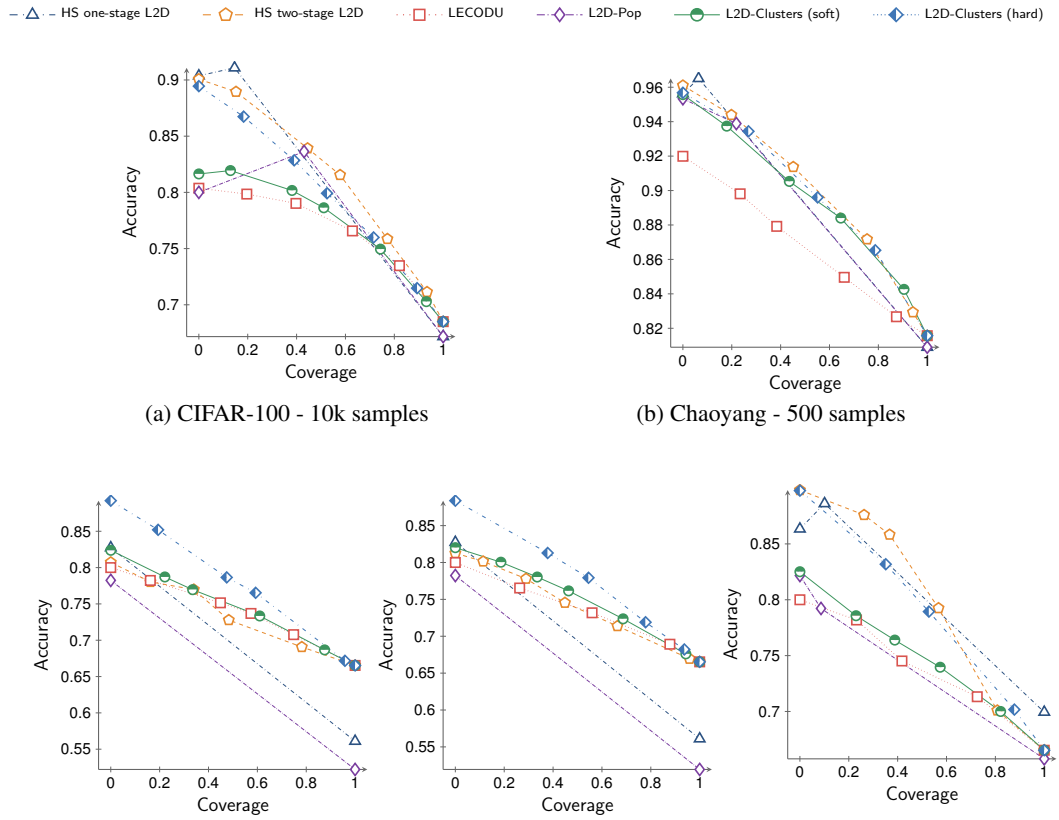
where: $\varepsilon_{\text{router}} > 0$ is the generalisation error of the router for any given $\widehat{\Omega}$, then:

$$
\begin{aligned}
\mathcal{R}(\widehat{g}, \widehat{\Omega}) - \mathcal{R}(g^*_{\Omega^*}, \Omega^*) &= \mathcal{R}(\widehat{g}, \widehat{\Omega}) - \mathcal{R}(g^*_{\widehat{\Omega}}, \widehat{\Omega}) + \mathcal{R}(g^*_{\widehat{\Omega}}, \widehat{\Omega}) - \mathcal{R}(g^*_{\Omega^*}, \Omega^*) \\
&\leq \varepsilon_{\text{router}} + [\min_{g\in\mathcal{G}} \mathcal{R}(g, \widehat{\Omega})] - \mathcal{R}(g^*_{\Omega^*}, \Omega^*) \\
&\leq \varepsilon_{\text{router}} + \mathcal{R}(g^*_{\Omega^*}, \widehat{\Omega}) - \mathcal{R}(g^*_{\Omega^*}, \Omega^*) \\
&\leq \varepsilon_{\text{router}} + \frac{2L}{s_{\min}} \sum_{k=1}^{K} \|\widehat{\Omega}_k - \Omega^*_k\|_1.
\end{aligned}
\tag{23}
$$

□

C ADDITIONAL L2D RESULTS

We provide additional results on L2D with a larger number of identified/context samples and show them in Fig. 4. In this setting, the conventional human-specific L2D methods (i.e., one-stage and two-stage) have a sufficient number of identified/context data to defer the decision to the most suitable experts (either human or classifier), and hence, upper-bound the performance of the remaining baselines and our method.



(a) CIFAR-100 - 10k samples

(b) Chaoyang - 500 samples

(c) dopanim - 100 samples

(d) dopanim - 200 samples

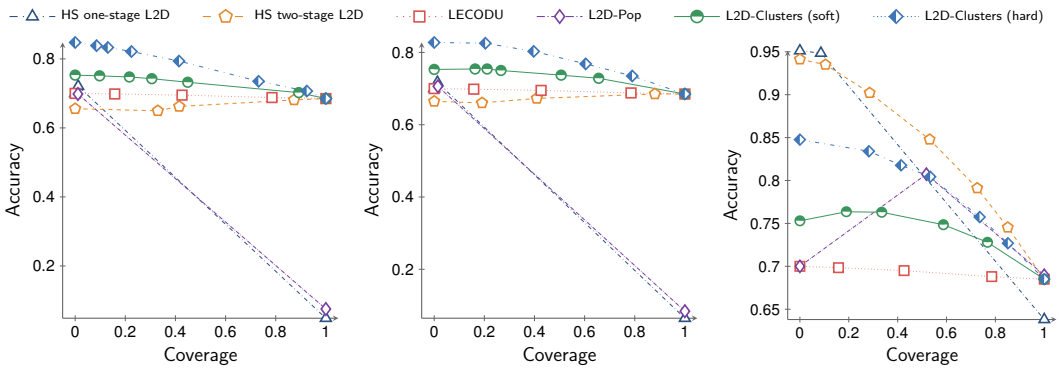
(e) dopanim - 4,497 samples

Figure 4: Additional results of coverage - accuracy curves on three datasets at a large number of identified/context samples.

D EXPERIMENT ON LARGE NUMBER OF EXPERTS

To evaluate our proposed framework, including both the LDA-based clustering and L2D-Clusters on a large number of experts, we simulate a total of 100 synthetic experts on CIFAR-100. In particular, a total of 10 sets of synthetic experts is simulated, where each set consists of 10 synthetic experts having the same error rate in the set $\{0.05, 0.1, 0.15, \ldots, 0.5\}$. After performing the clustering on different numbers of clusters shown in Fig. 9b, $K = 4$ clusters are selected to train an L2D-Clusters system. We follow a similar setting presented in Section 4.2 for the evaluation.

The results in Fig. 5 generally agree with our previous results evaluated on 6 synthetic experts. Specifically, our proposed L2D-Clusters method outperforms the baselines and state-of-the-art methods in the onboarding setting, where the context set size N' is small, such as $N' = 100$ as in Fig. 5a and $N' = 200$ as in Fig. 5b. In these cases, both the *soft* and *hard* variants of L2D-Clusters outperform the baselines and existing methods by large margins. The performance of both one-stage L2D and L2D-Pop is significantly degraded, in which the prediction accuracy of the classifier integrated into their model is below 10 percent. This is mainly due to their single-stage nature making the integrated classifier and the gating model overfit to the small number of context set. The two-stage L2D baseline performs better than the two-stage-like methods due to its modularity using a classifier trained on ground truth data, but its performance is even worse than the random selection LECODU method. However, when the context set size is large, such as $N' = 10,000$ as in Fig. 5c, there is a sufficient amount of data to identify the characteristics of each expert, and hence, the single-stage and two-stage L2D methods perform better than L2D-Clusters, at least better than L2D-Clusters soft, due to their human-specific nature.



(a) $N' = 100$ samples (b) $N' = 200$ samples (c) $N' = 10,000$ samples

Figure 5: The accuracy - coverage curves of different methods evaluated on different sizes of the context sets in the setting consisting of 100 synthetic experts simulated in CIFAR-100.

E Details of the parameter inference for the LDA model

E.1 Variational Expectation-Maximisation for LDA

Before going through the detailed derivation and analysis, we emphasise that our clustering is performed exclusively in the space of sets of annotations, denoted by $\mathcal{S}_i = \{(\mathbf{x}_{in}, \mathbf{t}_{in})\}_{n=1}^{N_i}$, and not in the feature space of the data instance $\mathbf{x}$ or the concatenated feature and annotation space $(\mathbf{x}, \mathbf{t})$. The resulting clusters therefore capture annotation similarity of human experts (e.g., agreement, disagreement, or bias), rather than latent classes or structural groupings of the data itself. Importantly, these clusters are not assumed to correspond to ground-truth labels, nor to any intrinsic structure of data instances.

The LDA-based method presented in this section and in Section 3.1 is a variant of the mini-batch based LDA in (Hoffman et al., 2010). Therefore, the proposed method also benefits from the convergence analysis in (Hoffman et al., 2010, Section 2.3), where the learning can be considered as a stochastic natural gradient algorithm (Sato, 2001).

To infer the parameters $(\theta_k)_{k=1}^K$ in Eq. (6), we follow the online learning approach for LDA in (Hoffman et al., 2010). In particular, the variational Expectation-Maximisation (EM) algorithm is employed to optimise a lower-bound of the objective function in (6). For simplicity, we consider one mini-batch, corresponding to one document, at a time, and therefore, omit the document index i from our analysis. In that case, the "complete-data" log-likelihood in Eq. (6) for $\mathcal{S}_i$ is defined as follows:

$$
\begin{aligned}
&Q\left(\mathbf{x}_{1:N_i}, \mathbf{t}_{1:N_i}, (\theta_k^{(\tau)})_{k=1}^K, (\theta_k)_{k=1}^K\right) \\
&= \mathbb{E}_{q(\mathbf{u}, \mathbf{z}_{1:N_i} | \mathbf{x}_{1:N_i}, \mathbf{t}_{1:N_i}, \alpha, (\theta_k^{(\tau)})_{k=1}^K)} \left[\ln \Pr(\mathbf{t}_{1:N_i}, \mathbf{z}_{1:N_i}, \mathbf{u} | \mathbf{x}_{1:N_i}, \alpha, (\theta_k)_{k=1}^K)\right] \\
&= \mathbb{E}_{q(\mathbf{u}, \mathbf{z}_{1:N_i} | \mathbf{x}_{1:N_i}, \mathbf{t}_{1:N_i}, \alpha, (\theta_k^{(\tau)})_{k=1}^K)} \left[\ln \Pr(\mathbf{t}_{1:N_i} | \mathbf{x}_{1:N_i}, \mathbf{z}_{1:N_i}, (\theta_k)_{k=1}^K) + \ln \Pr(\mathbf{z}_{1:N_i} | \mathbf{u})\right. \\
&\quad \left. + \ln \Pr(\mathbf{u} | \alpha)\right] \\
&= \mathbb{E}_{q(\mathbf{u}, \mathbf{z}_{1:N_i} | \mathbf{x}_{1:N_i}, \mathbf{t}_{1:N_i}, \alpha, (\theta_k^{(\tau)})_{k=1}^K)} \left[\textstyle\sum_{n=1}^{N_i} \ln \Pr(\mathbf{t}_n | \mathbf{x}_n, \mathbf{z}_n, (\theta_k)_{k=1}^K) + \ln \Pr(\mathbf{z}_n | \mathbf{u})\right. \\
&\quad \left. + \ln \Pr(\mathbf{u} | \alpha)\right], \qquad (24)
\end{aligned}
$$

where the superscript τ denotes the value at the τ-th iteration.

E-step Due to the complexity of the exact posteriors of the latent variables $\mathbf{u}$ and $\mathbf{z}$, we resort to the mean-field variational inference to approximate the posteriors. Firstly, the posterior of the two latent variables is assumed to be factorised as follows:

$$
\begin{aligned}
q(\mathbf{u}, \mathbf{z}_{1:N_i} | \mathbf{x}_{1:N_i}, \mathbf{t}_{1:N_i}, \alpha, (\theta_k^{(\tau)})_{k=1}^K) &= q(\mathbf{u}; \rho) \prod_{n=1}^{N_i} q(\mathbf{z}_n; \zeta_n) \\
&= \mathrm{Dir}(\mathbf{u}; \rho) \prod_{n=1}^{N_i} \mathrm{Categorical}(\mathbf{z}_n; \zeta_n), \qquad (25)
\end{aligned}
$$

where: $\rho \in (0, +\infty)^K$ is the parameter of the Dirichlet variational posterior, and $\zeta_n \in \Delta^{K-1}$ is the parameter of the categorical variational posterior for the n-th data point.

The approximate posterior of the two latent random variables $\mathbf{u}$ and $\mathbf{z}$ can be obtained via the mean-field variational inference as follows:

$$
\begin{aligned}
&(\rho^*, (\zeta_n^*)_{n=1}^{N_i}) \\
&= \underset{(\rho, (\zeta_n)_{n=1}^{N_i})}{\arg\min}\ \mathrm{KL}\left[q(\mathbf{u}, \mathbf{z}_{1:N_i} | \rho, (\zeta_n)_{n=1}^{N_i}) \,\middle\|\, \Pr(\mathbf{u}, \mathbf{z}_{1:N_i} | \mathbf{x}_{1:N_i}, \mathbf{t}_{1:N_i}, \alpha, (\theta_k^{(\tau)})_{k=1}^K)\right]. \qquad (26)
\end{aligned}
$$

Readers are referred to Appendix E.2 for the detailed derivation and optimisation of Eq. (26). The E-step is almost identical to the E-step in the latent Dirichlet allocation (Blei et al., 2003), except that the multinomial distributions, each representing a topic, are replaced by the categorical distributions, each being the output of a classifier (or an annotation style model). The parameters $(\rho^*, (\zeta_n^*)_{n=1}^{N_i})$ are then used to calculate the "complete-data" log-likelihood in Eq. (24).

M-step The "complete-data" log-likelihood estimated in Eq. (24) with ρ^* and $(\zeta_n^*)_{n=1}^{N_i}$ is then maximised w.r.t. $(\theta_k)_{k=1}^K$. This maximisation is performed by the gradient ascent algorithm with learning rate η, and the update of $(\theta_k)_{k=1}^K$ is given as follows:

$$\theta_k^{(\tau+1)} \leftarrow \theta_k^{(\tau)} + \eta \nabla_{\theta_k} \left[Q\left(\mathbf{x}_n, \mathbf{t}_n, (\theta_k^{(\tau)})_{k=1}^K, (\theta_k)_{k=1}^K \right) \right], \forall k \in \{1, \ldots, K\}. \tag{27}$$

E.2 Variational inference for the LDA-based modelling

In this section, we provide the detailed derivation to approximate the posterior of the latent random variables presented in Appendix E.1. The objective function of the variational inference is rewritten as follows:

$$(\rho^*, \{\zeta_n^*\}_{n=1}^{N_i}) = \underset{(\rho, \{\zeta_n\}_{n=1}^{N_i})}{\arg\min} \ \mathrm{KL}\left[q(\mathbf{u}, \mathbf{z}_{1:N_i} | \rho, \zeta_{1:N_i}) \| \Pr(\mathbf{u}, \mathbf{z}_{1:N_i} | \mathbf{x}_{1:N_i}, \mathbf{t}_{1:N_i}, \theta_{1:K}^{(\tau)}, \alpha) \right]. \tag{26}$$

The KL divergence on the right-hand side in Eq. (26) can be expanded as follows:

$$\begin{aligned}
&\mathrm{KL}\left[q(\mathbf{u}, \mathbf{z}_{1:N_i} | \rho, \zeta_{1:N_i}) \| \Pr(\mathbf{u}, \mathbf{z}_{1:N_i} | \mathbf{x}_{1:N_i}, \mathbf{t}_{1:N_i}, \theta_{1:K}^{(\tau)}, \alpha) \right] \\
&= \mathbb{E}_{q(\mathbf{u}, \mathbf{z}_{1:N_i} | \rho, \zeta_{1:N_i})} \left[\ln q(\mathbf{u}, \mathbf{z}_{1:N_i} | \rho, \zeta_{1:N_i}) - \ln \Pr(\mathbf{u}, \mathbf{z}_{1:N_i} | \mathbf{x}_{1:N_i}, \mathbf{t}_{1:N_i}, \theta_{1:K}^{(\tau)}, \alpha) \right] \\
&= \mathbb{E}_{q(\mathbf{u}|\rho)} \left[\ln q(\mathbf{u}|\rho) \right] \\
&\quad + \sum_{n=1}^{N_i} \mathbb{E}_{q(\mathbf{z}_n|\zeta_n)} \left[\ln q(\mathbf{z}_n|\zeta_n) \right] \\
&\quad - \sum_{n=1}^{N_i} \mathbb{E}_{q(\mathbf{z}_n|\zeta_n)} \left[\ln \Pr(\mathbf{t}_n | \mathbf{x}_n, \theta_{\mathbf{z}_n}^{(\tau)}) \right] \\
&\quad - \mathbb{E}_{q(\mathbf{u}|\rho)} \left[\sum_{n=1}^{N_i} \mathbb{E}_{q(\mathbf{z}_n|\zeta_n)} \left[\ln \Pr(\mathbf{z}_n | \mathbf{u}) \right] \right] \\
&\quad - \mathbb{E}_{q(\mathbf{u}|\rho)} \left[\ln \Pr(\mathbf{u}|\alpha) \right] + \text{const.} \\
&= \ln \Gamma\left(\sum_{k=1}^K \rho_k \right) - \sum_{k=1}^K \ln \Gamma(\rho_k) - \sum_{k=1}^K (\rho_k - 1) \left[\psi\left(\sum_{j=1}^K \rho_j \right) - \psi(\rho_k) \right] \\
&\quad + \sum_{n=1}^{N_i} \sum_{k=1}^K \zeta_{nk} \ln \zeta_{nk} \\
&\quad - \sum_{n=1}^{N_i} \sum_{k=1}^K \zeta_{nk} \times \left[\mathbf{t}_n^\top \ln h\left(\mathbf{x}; \theta_k^{(\tau)} \right) \right] \\
&\quad - \sum_{n=1}^{N_i} \sum_{k=1}^K \zeta_{nk} \left[\psi(\rho_k) - \psi\left(\sum_{j=1}^K \rho_j \right) \right] \\
&\quad + \sum_{k=1}^K \ln \Gamma\left(\alpha_k \right) - \ln \Gamma\left(\sum_{k=1}^K \alpha_k \right) - \sum_{k=1}^K (\alpha_k - 1) \left[\psi(\rho_k) - \psi\left(\sum_{j=1}^K \rho_j \right) \right],
\end{aligned} \tag{28}$$

where: $\Gamma(.)$ is the gamma function and $\psi(.)$ is the digamma function.

E.2.1 Variational inference for the parameter of categorical distribution

We minimise the variational-free energy in Eq. (28) w.r.t. each ζ_n. Note that ζ_n is a probability vector, meaning that:

$$\sum_{k=1}^K \zeta_{nk} = 1. \tag{29}$$

The Lagrangian for ζ_{nk} can then be written as follows:

$$\mathsf{L}_{\zeta_{nk}} = \zeta_{nk} \ln \zeta_{nk} - \zeta_{nk} \mathbf{t}_n^\top \ln h\left(\mathbf{x}; \theta_k^{(\tau)}\right) - \zeta_{nk}\left[\psi(\rho_k) - \psi\left(\sum_{j=1}^{K} \rho_j\right)\right] + \lambda_n\left(\sum_{j=1}^{K} \zeta_{nj} - 1\right). \tag{30}$$

Taking the derivative w.r.t. ζ_{nk} gives the following expression:

$$\frac{\mathrm{d}\mathsf{L}_{\zeta_{nk}}}{\mathrm{d}\zeta_{nk}} = \ln \zeta_{nk} + 1 - \mathbf{t}_n^\top \ln h\left(\mathbf{x}; \theta_k^{(\tau)}\right) - \psi(\rho_k) + \psi\left(\sum_{j=1}^{K} \rho_j\right) + \lambda_n. \tag{31}$$

Setting the derivative to zero yields the optimal solution as follows:

$$\zeta_{nk} \propto \exp\left[\mathbf{t}_n^\top \ln h\left(\mathbf{x}; \theta_k^{(\tau)}\right) + \psi(\rho_k) - \psi\left(\sum_{j=1}^{K} \rho_j\right)\right]. \tag{32}$$

E.2.2 Variational inference for the parameter of Dirichlet distribution

The variational-free energy for ρ_j can be written as follows:

$$\begin{aligned}
\mathsf{L}_\rho = {} & \ln \Gamma\left(\sum_{j=1}^{K} \rho_j\right) - \sum_{k=1}^{K} \ln \Gamma(\rho_k) - \sum_{k=1}^{K} (\rho_k - 1)\left[\psi\left(\sum_{j=1}^{K} \rho_j\right) - \psi(\rho_k)\right] \\
& - \sum_{n=1}^{N_i} \sum_{k=1}^{K} \zeta_{nk}\left[\psi(\rho_k) - \psi\left(\sum_{j=1}^{K} \rho_j\right)\right] \\
& + \sum_{k=1}^{K} (\alpha_k - 1)\left[\psi\left(\sum_{j=1}^{K} \rho_j\right) - \psi(\rho_k)\right] \\
= {} & \sum_{k=1}^{K}\left[\psi\left(\sum_{j=1}^{K} \rho_j\right) - \psi(\rho_k)\right]\left[\alpha_k - \rho_k + \sum_{n=1}^{N_i} \zeta_{nk}\right] \\
& + \ln \Gamma\left(\sum_{j=1}^{K} \rho_j\right) - \sum_{k=1}^{K} \ln \Gamma(\rho_k).
\end{aligned} \tag{33}$$

Taking the derivative with respect to ρ_k gives:

$$\frac{\mathrm{d}\mathsf{L}_\rho}{\mathrm{d}\rho_k} = \left[\psi'\left(\sum_{j=1}^{K} \rho_j\right) - \psi'(\rho_k)\right]\left[\alpha_k - \rho_k + \sum_{n=1}^{N_i} \zeta_{nk}\right], \tag{34}$$

where: $\psi'(.)$ is the trigamma function, which is the derivative of the digamma function.

Setting the derivative to zero yields the optimal solution as follows:

$$\rho_k = \alpha_k + \sum_{n=1}^{N_i} \zeta_{nk}. \tag{35}$$

F Error analysis of the test-time cluster assignment

We analyse the error incurred when inferring the cluster membership of a test-time human expert m from the identified set $\mathcal{I}_m = \{(\mathbf{x}_n, \mathbf{t}_n)\}_{n=1}^{N_m}$. Throughout this section, the parameters $\theta^*_{1:K}$ are already obtained through training by maximising the log-likelihood in (6), and are held fixed. The quantity of interest is the variational Dirichlet parameter $\rho^{(m)}$ of Eq. (35), since the cluster-assignment matrix fed into the downstream L2D system is, by Eq. (8), the column-wise normalisation of $\rho^{(m)}$.

Following Section 3.2, the identified set is treated as a single document of N_m words. The parameters ζ_{mn} and $\rho^{(m)}$ are inferred similar to the training case in Eqs. (32) and (35), which can be written as follows:

$$\zeta_{nk}^{(m)} \propto \exp\left\{\mathbb{E}_{q(\mathbf{u}|\rho^{(m)})}\left[\ln u_k\right] + \ln h_{\mathbf{t}_n}(\mathbf{x}_n; \theta_k^*)\right\}, \tag{36}$$

$$\rho_k^{(m)} = \alpha_k + \sum_{n=1}^{N_m} \zeta_{nk}, \tag{37}$$

The two equations above are repeatedly updated until convergence.

F.1 Proof for Proposition 2

For convenience, we restate the proposition here.

Proposition 2. *The cluster membership inferred from $\rho^{(m)}$ satisfies*

$$\|\hat{\omega}^{(m)} - \omega^{(m)}\|_1 \leq \min\left(2, \sqrt{2\,\mathrm{KL}[q(\mathbf{u}, \mathbf{z}|\rho^{(m)}, \zeta^{(m)}_{1:N_m})\|\Pr(\mathbf{u}, \mathbf{z}|\mathcal{I}_m, \alpha, (\theta_k^*)_{k=1}^K)]}\right). \tag{9}$$

Proof. Both assignment vectors are means of $\mathbf{u}$ over the simplex Δ^{K-1}, the estimate $\hat{\omega}^{(m)}$ under the variational posterior $q(\mathbf{u}|\rho^{(m)})$ and the reference $\omega^{(m)}$ under the exact one $\Pr(\mathbf{u}|\mathcal{I}_m, \alpha, (\theta_k^*)_{k=1}^K)$. The L_1 error of the cluster assignment probability vector can be written as follows:

$$\begin{aligned}
\|\hat{\omega}^{(m)} - \omega^{(m)}\|_1 &= \left\| \int \mathbf{u}\, q(\mathbf{u}|\rho^{(m)}) \mathrm{d}\mathbf{u} - \int \mathbf{u}\, \Pr(\mathbf{u}|\mathcal{I}_m, \alpha, (\theta_k^*)_{k=1}^K) \mathrm{d}\mathbf{u} \right\|_1 \\
&= \left\| \int \mathbf{u} \left[q(\mathbf{u}|\rho^{(m)}) - \Pr(\mathbf{u}|\mathcal{I}_m, \alpha, (\theta_k^*)_{k=1}^K) \right] \mathrm{d}\mathbf{u} \right\|_1 \\
&= \sum_{k=1}^K \left| \int \mathbf{u}_k \left[q(\mathbf{u}|\rho^{(m)}) - \Pr(\mathbf{u}|\mathcal{I}_m, \alpha, (\theta_k^*)_{k=1}^K) \right] \mathrm{d}\mathbf{u} \right| \\
&\leq \sum_{k=1}^K \int |\mathbf{u}_k| \left| q(\mathbf{u}|\rho^{(m)}) - \Pr(\mathbf{u}|\mathcal{I}_m, \alpha, (\theta_k^*)_{k=1}^K) \right| \mathrm{d}\mathbf{u} \\
&= \sum_{k=1}^K \int \mathbf{u}_k \left| q(\mathbf{u}|\rho^{(m)}) - \Pr(\mathbf{u}|\mathcal{I}_m, \alpha, (\theta_k^*)_{k=1}^K) \right| \mathrm{d}\mathbf{u} \quad (\text{due to: } \mathbf{u}_k \geq 0) \\
&= \int \left(\sum_{k=1}^K \mathbf{u}_k \right) \left| q(\mathbf{u}|\rho^{(m)}) - \Pr(\mathbf{u}|\mathcal{I}_m, \alpha, (\theta_k^*)_{k=1}^K) \right| \mathrm{d}\mathbf{u} \\
&= \int \left| q(\mathbf{u}|\rho^{(m)}) - \Pr(\mathbf{u}|\mathcal{I}_m, \alpha, (\theta_k^*)_{k=1}^K) \right| \mathrm{d}\mathbf{u} \quad (\sum_{k=1}^K u_k = 1) \\
&= 2\,\mathrm{TV}[q(\mathbf{u}|\rho^{(m)}), \Pr(\mathbf{u}|\mathcal{I}_m, \alpha, (\theta_k^*)_{k=1}^K)],
\end{aligned} \tag{38}$$

where $\mathrm{TV}[\cdot, \cdot]$ is the total variation distance between two distributions.

Applying Pinsker's inequality, $\mathrm{TV}(q, p) \leq \sqrt{\mathrm{KL}[q\|p]/2}$, to Eq. (38) gives

$$\|\hat{\omega}^{(m)} - \omega^{(m)}\|_1 \leq \sqrt{2\,\mathrm{KL}\left[q(\mathbf{u}|\rho^{(m)})\|\Pr(\mathbf{u}|\mathcal{I}_m, \alpha, (\theta_k^*)_{k=1}^K)\right]}. \tag{39}$$

The divergence in Eq. (39) is between the marginals of $\mathbf{u}$ alone, whereas the inference of Eq. (7) operates on the joint distribution of $\mathbf{u}$ and $\mathbf{z}$. The two are related by the chain rule for the Kullback-Leibler divergence,

$$\begin{aligned}
&\mathrm{KL}\left[q(\mathbf{u},\mathbf{z}|\rho^{(m)},\zeta^{(m)}_{1:N_m})\|\Pr(\mathbf{u},\mathbf{z}|\mathcal{I}_m,\alpha,(\theta^*_k)_{k=1}^K)\right]\\
&=\mathrm{KL}\left[q(\mathbf{u}|\rho^{(m)})\|\Pr(\mathbf{u}|\mathcal{I}_m,\alpha,(\theta^*_k)_{k=1}^K)\right]+\mathbb{E}_{q(\mathbf{u}|\rho^{(m)})}\left[\mathrm{KL}\left[q(\mathbf{z}|\zeta^{(m)}_{1:N_m})\|\Pr(\mathbf{z}|\mathbf{u},\mathcal{I}_m)\right]\right]\\
&\geq\mathrm{KL}\left[q(\mathbf{u}|\rho^{(m)})\|\Pr(\mathbf{u}|\mathcal{I}_m,\alpha,(\theta^*_k)_{k=1}^K)\right], \qquad (40)
\end{aligned}$$

where the inequality holds because the discarded term is an expectation of a KL divergence, and hence non-negative. The left-hand side of Eq. (40) is precisely the objective minimised in Eq. (7), evaluated at the returned parameters $\rho^{(m)}$ and $\zeta^{(m)}_{1:N_m}$.

Finally, $\hat{\omega}^{(m)}$ and $\omega^{(m)}$ are both probability vectors, so their L_1 distance never exceeds 2. Combining this with Eqs. (39) and (40) gives the claim:

$$\|\hat{\omega}^{(m)}-\omega^{(m)}\|_1\leq\min\left(2,\sqrt{2\,\mathrm{KL}[q(\mathbf{u},\mathbf{z}|\rho^{(m)},\zeta^{(m)}_{1:N_m})\|\Pr(\mathbf{u},\mathbf{z}|\mathcal{I}_m,\alpha,(\theta^*_k)_{k=1}^K)]}\right). \qquad (9)$$

□

F.2 Proof for Corollary 1

For convenience, we restate the corollary here.

Corollary 1. *Let* $\mathrm{KL}_{\max}=\max_{m\in\{1,\dots,M\}}\mathrm{KL}[q(\mathbf{u},\mathbf{z}|\rho^{(m)},\zeta^{(m)}_{1:N_m})\|\Pr(\mathbf{u},\mathbf{z}|\mathcal{I}_m,\alpha,(\theta^*_k)_{k=1}^K)]$ *be the worst-case variational gap over the* M *experts. The excess risk of L2D-Clusters is bounded as:*

$$\mathcal{R}(\widehat{g},\widehat{\Omega})-\mathcal{R}(g^*_{\Omega^*},\Omega^*)\leq\tfrac{2LM}{s_{\min}}\times\min\left(2,\sqrt{2\,\mathrm{KL}_{\max}}\right)+\varepsilon_{\text{router}}.$$

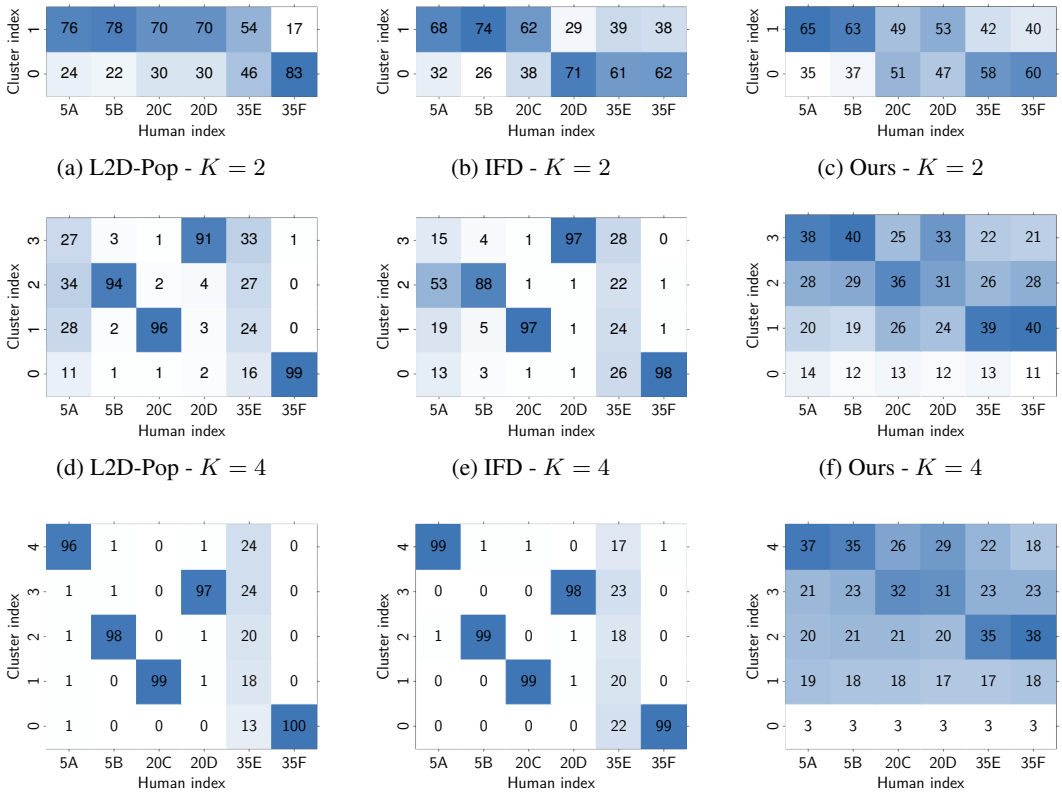
Proof. The reference cluster assignment matrix Ω^* is defined to be the matrix of exact-posterior assignments, whose m-th column is $\omega^{(m)}$, is therefore admissible whenever its clusters retain mass $s_{\min}$. The cluster-assignment matrix $\widehat{\Omega}$ is formed by stacking the membership vectors $\hat{\omega}^{(m)}$ column-wise, as described in Section 3.2. Consequently, the row-wise L_1 perturbation appearing in Proposition 1 is the entry-wise absolute deviation of the matrix, which can equally be summed column-wise. Applying Eq. (9) to each expert and bounding every variational gap by the worst case $\mathrm{KL}_{\max}$,

$$\sum_{k=1}^K\left\|\widehat{\Omega}_k-\Omega^*_k\right\|_1=\sum_{m=1}^M\left\|\hat{\omega}^{(m)}-\omega^{(m)}\right\|_1\leq M\min\left(2,\sqrt{2\,\mathrm{KL}_{\max}}\right). \qquad (41)$$

Substituting Eq. (41) into the excess-risk bound of Proposition 1 completes the proof. □

G ADDITIONAL CLUSTERING RESULTS

We provide additional clustering results on CIFAR-100 with $K \in \{4, 5\}$ clusters in Fig. 6. In general, L2D-Pop and IFD assign every synthetic human expert into the available clusters with high confidence, although the assignment may not agree with the simulation. In contrast, our results in Figs. 6f and 6i consistently assign synthetic experts 5A and 5B into one group, and 35E and 35F into another group, while having high uncertainty assigning the remaining pair of synthetic experts 20C and 20D. Furthermore, the proposed LDA-based clustering method does not require every cluster to be occupied, but only the clusters supported by the data.



(a) L2D-Pop - $K = 2$ (b) IFD - $K = 2$ (c) Ours - $K = 2$

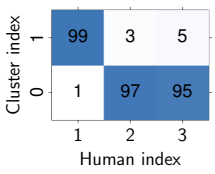
(d) L2D-Pop - $K = 4$ (e) IFD - $K = 4$ (f) Ours - $K = 4$

(g) L2D-Pop - $K = 5$ (h) IFD - $K = 5$ (i) Ours - $K = 5$

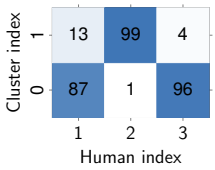
Figure 6: Additional clustering results on CIFAR-100 with the context/identified set size of $N' = 100$ samples, in which the numbers in each column represent the probability of cluster assignment for each synthetic expert. Note that the colour in each heatmap is relatively scaled to enhance the visualisation.

H CLUSTERING ON REAL-WORLD DATASETS

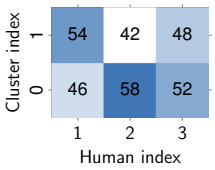
We show the clustering result on Chaoyang in Fig. 7. Both baseline methods, L2D-Pop and IFD, implicitly prioritise overall accuracy and produce near-hard assignments (e.g., pathologists 1 and 3 co-cluster due to similar accuracies of 91% and 99% vs 88% from pathologist 2). Our method yields softer memberships that reflect both agreement and disagreement structure: pathologists 1 and 2 exhibit the smallest pairwise agreement ($\approx$ 79%), while pathologist 3 shares comparable agreement with both 1 and 2 (91% and 87%, respectively). That is why pathologists 1 and 2 are grouped into two separate clusters with slightly higher probability, while pathologist 3 is most likely in both clusters.



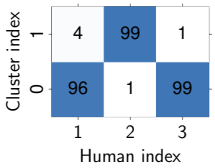
(a) L2D-Pop - $N' = 50$



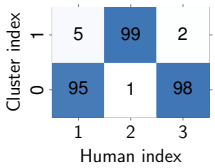
(b) IFD - $N' = 50$



(c) Ours - $N' = 50$



(d) L2D-Pop - $N' = 100$



(e) IFD - $N' = 100$



(f) Ours - $N' = 100$

Figure 7: Clustering results on the dataset Chaoyang at different numbers of context (i.e., L2D-Pop and IFD) and identified sets (ours): *(top row)* $N' = 50$ and *(bottom row)* $N' = 100$ samples.

I SENSITIVITY ANALYSIS ON THE NUMBER OF CLUSTERS K

Sensitivity analysis on the number of clusters is performed on CIFAR-100 with 100 experts and identified set size of 100 samples and reported as the area under coverage-accuracy curve (AUC-AC) for different values of K in Fig. 8. Increasing the number of clusters K can indeed allow the model to capture finer-grained distinctions in annotation agreement within the anonymously annotated dataset. However, setting K excessively high introduces redundant clusters whose probabilities remain near zero across all annotation sets, effectively making them "unused". Empirically, we observe that the dataset exhibits an intrinsic cluster structure with $K_{\text{intrinsic}} = 4$. Beyond this point, additional clusters do not contribute meaningful information, leading to a plateau in performance.

Despite that, fixing $K = K_{\text{intrinsic}}$ is not always optimal, as it may risk overfitting to the training data. To address this, our paper adopts a principled heuristic based on the *rate of perplexity change* (Zhao et al., 2015) evaluated on a held-out set (details in Appendix K). This metric provides a more robust criterion for selecting K. Importantly, our experiments show that performance remains stable across a range of K, indicating that the model is relatively insensitive to this hyperparameter.

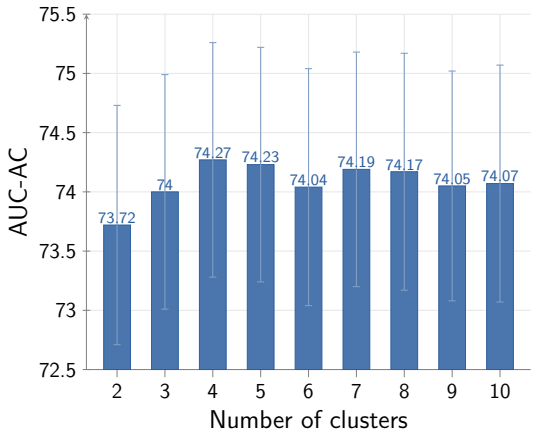


Figure 8: Sensitivity analysis on the number of clusters K in which the AUC-AC results are stable at various K.

J SENSITIVITY ANALYSIS ON THE DIRICHLET PRIOR

Sensitivity analysis on Dirichlet prior is also performed to quantify clustering stability by computing the average pairwise cosine dissimilarity of the cluster assignment distributions defined in Eq. (8). This metric captures how dissimilar the cluster membership probabilities are between every pair of human experts, in which a larger value represents a more diverse clustering. In addition, we report the AUC-AC score in Table 3 (see Appendix J), which reflects the overall system performance across coverage-accuracy trade-offs. Further increasing α leads to a plateau in both the pairwise cosine similarity and the performance (i.e., AUC-AC). This occurs because α governs the prior cluster distribution $\Pr(\mathbf{u}|\alpha)$, whereas the posterior of cluster assignment for the human expert indexed by m (see Section 3.2), denoted as $\Pr(\mathbf{u}|\alpha, \mathcal{I}_m)$, depends on the observed identified set $\mathcal{I}_m$. Thus, while a larger α promotes uniformity of cluster distribution, the posterior $\Pr(\mathbf{u}|\alpha, \mathcal{I}_m)$ for the human expert m does not need to be uniform. If an expert's annotations strongly favour certain clusters, the posterior will remain skewed toward those clusters despite the prior bias, which results in the plateau in both the pairwise cosine dissimilarity and AUC-AC.

α_0	**Pair-wise cosine dissimilarity** (↑)	**AUC-AC** ($\times 100 \uparrow$)
0.1	0.0002	72.90 ± 2.00
0.9	0.0009	73.10 ± 1.77
2.0	0.0377	74.98 ± 1.32
10	0.0406	75.07 ± 1.06
50	0.0422	75.04 ± 1.25

Table 3: The area under coverage-accuracy curves of our two-stage method (including both clustering and L2D-Clusters) at various values of the hyper-parameter representing the Dirichlet prior α_0, in which a larger value of α_0 promotes a more diverse clustering (higher cosine dissimilarity of cluster assignment probability pair).

K DETERMINE THE NUMBER OF CLUSTERS

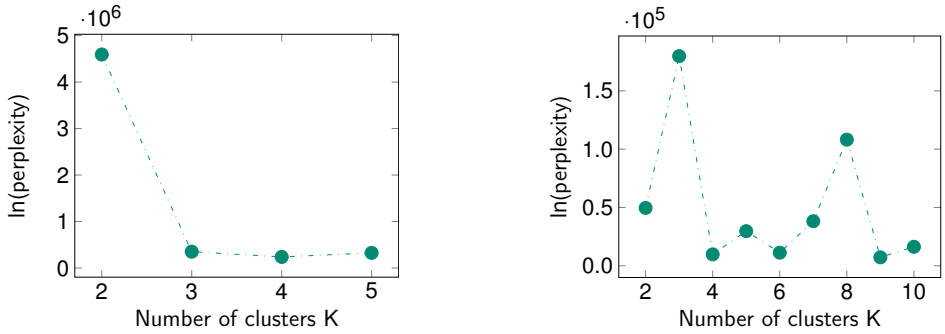
Summary The number of clusters can be selected through various approaches, such as non-parametric learning with *Dirichlet process* (Blei & Jordan, 2004) or hyper-parameter optimisation. For simplicity, we follow the latter approach by evaluating the perplexity with different number of clusters K on a hold-out set of anonymously-annotated data. We then select K that minimises perplexity. However, perplexity is known to exhibit a tendency toward overfitting, whereby increasing K often leads to smaller perplexity values regardless of model quality (Chang et al., 2009). To mitigate this issue, we adopt a heuristic metric known as the *rate of perplexity change* (Zhao et al., 2015), which identifies the smallest value of K beyond which further relative reductions in perplexity become marginal. For example, in the simulation on CIFAR-100 with six synthetic human experts, the optimal number of clusters is determined to be $K = 3$, corresponding to the point of maximum rate of perplexity change on the hold-out set (see Fig. 9a).

The perplexity in our case is modified accordingly, in which it is the likelihood of the hold-out data:

$$\text{perplexity}(\mathcal{S}_{\text{val}}) = \exp(-\frac{\sum_{n=1}^{|\mathcal{S}_{\text{val}}|} \ln \Pr(\mathbf{t}_n|\mathbf{x}_n, \alpha, (\theta_k^*)_{k=1}^K)}{\sum_{n=1}^{|\mathcal{S}_{\text{val}}|} \mathbf{1}^\top \mathbf{t}_n}). \tag{42}$$

The smaller the perplexity, the better fit our model is to the data.

We train several LDA-based models with different number of clusters K on the dataset CIFAR-100 and plot the perplexity evaluated on a hold-out validation set in Fig. 9.



(a) CIFAR-100 - 6 synthetic experts

(b) CIFAR-100 - 100 synthetic experts

Figure 9: Perplexity (smaller is better) on the hold-out validation set of the dataset CIFAR-100.

To select the optimal number of clusters K, we employ the heuristic metric *rate of perplexity change* (RPC) in (Zhao et al., 2015) defined as follows:

$$\text{RPC}_i = |\frac{P_i - P_{i-1}}{L_i - L_{i-1}}|, \tag{43}$$

where: P_i is the perplexity evaluated with L_i clusters.

The optimal number of clusters is selected as the smallest number of clusters that give the highest relative improvement:

$$K^* = \min \operatorname*{argmax}_i \text{RPC}_i \,. \tag{44}$$

Hence, the optimal number of clusters for the setting of CIFAR-100 with 6 and 100 synthetic experts are $K = 3$, and $K = 4$, respectively.

L COMPUTATION COST

We report the training time of both stages: LDA and L2D-Cluster. These are evaluated on an Nvidia A6000 GPU. Note that the running time of LDA mostly depends on the number of clusters K (i.e., number of models $h(.;\theta_k)$), and does not vary much w.r.t. the number of experts M because of batching (i.e., parallelism) in GPU.

Table 4: Computation time (GPU-h) for LDA-based expert clustering and L2D-Clusters.

Dataset	**Number of clusters**	**LDA**
CIFAR-100 (M = 100)	3	1.3
	4	1.4
	5	1.4
	6	1.6
	7	1.9
	8	2.1
	9	2.3
	10	2.6
dopanim	2	0.5
	3	0.98
Chaoyang	2	1.1

For the L2D-Clusters, running time is modest and often less than 30 minutes due to the small size of identified sets used (i.e., 50, 100, 200 and 10,000 samples).

M ALGORITHM TO CLUSTER HUMANS ON ANONYMOUS ANNOTATION DATA

We provide the details of the two procedures proposed in the paper:

1. Learning the clusterer using anonymously annotated data, and
2. Inferring the probability of cluster membership for each human expert based on his or her identified set.

Refer to Algorithm 1 for more details.

Algorithm 1 Learning to defer with anonymous annotation data

procedure MODEL-ANONYMOUS-ANNOTATION-DATA($\mathcal{S} = \{(\mathbf{x}_n, \mathbf{t}_n)\}_{n=1}^N, \alpha, K$)
  ▷ $\mathbf{x}_n$: *input sample* ◁
  ▷ $\mathbf{t}_n$: *multinomial vector of anonymous annotations* ◁
  ▷ α: *the concentration parameter of the Dirichlet prior* ◁
  ▷ K: *number of clusters* ◁
  initialise K parameters $(\theta_k^{(0)})_{k=1}^K$
  for $i \in \{0, \dots, N-1\}$ **do** ▷ *a total of N iterations*
    sample a subset S_i consisting of N_i samples from the set of anonymous annotations
    obtained variational parameters: $(\rho_i^*, (\zeta_{ij}^*)_{j=1}^{N_i}) \leftarrow \mathrm{VI}(\mathbf{t}_i, \mathbf{x}_i, \alpha, (\theta_k^{(i)})_{k=1}^K)$ ▷ *Eq. (26)*
    E-step: calculate $Q\left(\mathbf{x}_n, \mathbf{t}_n, \rho_i^*, (\zeta_{ij}^*)_{j=1}^{N_i}, \alpha, (\theta_k^{(\tau)})_{k=1}^K, (\theta_k)_{k=1}^K\right)$ ▷ *Eq. (24)*
    M-step: $(\theta_k^{(\tau+1)})_{k=1}^K \leftarrow \mathrm{argmax}_{(\theta_k)_{k=1}^K} Q\left(\mathbf{x}_n, \mathbf{t}_n, \rho_i^*, (\zeta_{ij}^*)_{j=1}^{N_i}, \alpha, (\theta_k^{(\tau)})_{k=1}^K, (\theta_k)_{k=1}^K\right)$ ▷ *Eq. (27)*
  return $(\theta_k^{(T)})_{k=1}^K$

procedure INFER-CLUSTER-ASSIGNMENT-PROBABILITY($\mathcal{I}_m, \alpha, (\theta_k^*)_{k=1}^K$)
  ▷ $\mathcal{I}_m$: *a small set of data annotated by a human expert indexed by m* ◁
  ▷ α: *a hyperparameter representing the parameter of the Dirichlet prior* ◁
  ▷ θ_k^*: *learnt parameter of each "cluster" component* ◁
  obtain variational parameters: $(\rho_m^*, (\zeta_{mj}^*)_{j=1}^{N_m}) \leftarrow \mathrm{VI}(\mathbf{t}_m, \mathbf{x}_m, \alpha, (\theta_k^*)_{k=1}^K)$ ▷ *Eq. (26)*
  calculate cluster-assigned probability: $\boldsymbol{\omega}^{(m)} \leftarrow {}^{\rho_m^*}/_{\mathbf{1}^\top \rho_m^*}$ ▷ *Eq. (8)*
  return $\boldsymbol{\omega}$

N Datasets and hyper-parameters

N.1 Datasets

CIFAR-100 consists of 50,000 training and 10,000 testing images which are grouped into 20 superclasses. Each of these superclasses consists of 5 sub-classes, making up a total of 100 classes for the whole dataset. We follow the approach of (Hemmer et al., 2023; Nguyen et al., 2025) to synthetically generate three pairs of experts, each exhibiting instance-dependent label noise (Xia et al., 2020). In particular, each synthetic expert in one pair would mis-classify with an error of 5, 20 and 35 percent. The reason for these error rates is to ensure that synthetic human performance is still higher than a classifier (the accuracy of the classifier is approximately 67 percent). And because the original CIFAR-100 testing set has about 10 percent of duplications or almost-duplications to the ones in the training set, we employ ciFAIR-100 (Barz & Denzler, 2020), which replaces those duplicated images by ones that belong to the same classes sampled from the ImageNet dataset.

dopanim is a small subset of the notorious dataset iNaturalist. The dataset consists of 15,750 animal images of 15 classes, grouped into four groups of doppelganger animals and collected together with ground truth labels (Herde et al., 2024). The dataset is split into a training subset consisting of 10,500 images, and a validation subset consisting of the remaining images. We follow a similar strategy in CIFAR-100 to simulate three pairs of human annotators, each pair classifies with a mistake of 5, 20 and 35 percent, respectively.

Chaoyang is a pathology image dataset of colon patches obtained from the Chaoyang hospital in China (Zhu et al., 2022). It consists of 4,021 samples for training and 2,139 samples for testing. Three professional pathologists provide annotations over four categories: normal, serrated, adenocarcinoma, and adenoma with the ground truth labels obtained via the majority vote of the three annotations. The performance of the three pathologists is 91, 87 and 99 percent accurate when considering majority vote as the ground truth.

N.2 Experiment setting

In general, we follow the same experiment settings in previous L2D studies (Nguyen et al., 2025) to evaluate our proposed methods.

Anonymous and identified subsets For each dataset, we randomly split the training set into training and validation. In particular, in CIFAR-100, we use 40,000 samples for training, while preserving 10,000 samples as either the context set for L2D-Pop and EA-L2D or identified set for our method. In Chaoyang, we randomly sample 500 samples without replacement to be the context/identified set. The remaining samples are used as anonymously-annotated data to train the LDA-based model. During the evaluation, we randomly select N' samples out of those hold-out samples as annotator-identified data to (i) infer the cluster probability for each expert, and (ii) perform L2D-Clusters.

Models Preact-Resnet-18 is used on the dataset CIFAR-100, while the conventional Resnet-18 is used on the datasets dopanim and Chaoyang.

Hyper-parameters For batch size, we use 128 samples on CIFAR-100, and 32 samples on Chaoyang and dopanim and optimise via stochastic gradient descent with a momentum of 0.9 and a learning rate of 0.001 decayed through a cosine annealing for 150 epochs, resulting in approximately 1.2 GPU-hour when running on one NVIDIA GPU A6000. For the LDA-based clustering model, due to the large overlap between the clusters of human experts (most humans annotate similarly except few cases), the Dirichlet concentration prior α must be large to avoid collapsing, in which any expert will be assigned to the same cluster. In the experiment, we set $\alpha_0 = 50$ for all datasets. For the truncated Gaussian prior, we set $\mu_\alpha = 5$, $\alpha_{\min} = 2$, and $\alpha_{\max} = +\infty$ across all experiments. Excessively large values of β shrink every expert-specific parameter towards μ_α, reducing cluster differentiation while tightening the bound in Proposition 2. We therefore use the small fixed value $\beta = 10^{-4}$ in all experiments to retain expert-specific variation. This value is selected using the anonymous validation split.

O EXTENDED RELATED WORK

Clustering crowdsource annotators methods aim to group human annotators based on the similarity of their annotation patterns. Early approaches in this domain, such as *crowd-parting* (Kairam & Heer, 2016), simply measured the agreement and disagreement on shared annotated samples to quantify the similarity of annotators. Subsequent methods have employed more sophisticated statistical techniques, including Cohen's kappa and Krippendorff's alpha coefficients, to assess inter-annotator agreement and correlation (Wich et al., 2020; Basile, 2020; Lo & Basile, 2023). These methods, whether based on simple agreement counts or statistical coefficients, rely on overlapping samples, restricting their applicability in more general scenarios, such as distinct, non-overlapping subsets of data. More sophisticated modelling approaches based on probabilistic graphical models have also been proposed to address more complex annotation settings and model annotator reliability or bias (Nguyen et al., 2019; Traganitis & Giannakis, 2022). These models, however, assume a fixed set of experts, typically requiring retraining or complex inference procedures upon changes to the expert pool, which is incompatible with the setting of real-world dynamics we consider in this paper. Another line of research focuses on directly learning representations for individual human experts, which can then be used for clustering (Deng et al., 2023; Tailor et al., 2024). A key limitation of these representation-learning methods is the need for annotator-identified training data. In contrast to these existing approaches, our clustering method, detailed in Section 3.1, operates independently of the annotator-identified data during training. Furthermore, it can effectively utilise any set of annotated samples, irrespective of overlapping set of samples or the number of annotators per sample. The trained model can subsequently be used to cluster newly on-boarded annotators directly, without the need for re-training. Such capabilities of our proposed method offer enhanced flexibility and broader applicability, particularly in scenarios involving anonymously annotated datasets.

Consensus labels from crowd is another related research topic aiming to elicit ground truth labels from manual annotations. In particular, these studies require access to identified annotations of human experts coupled either with probabilistic modelling (Dawid & Skene, 1979) or with pre-trained classifiers to produce consensus labels (Showalter et al., 2024; Kelly et al., 2025). This is distinctively different from our study, in which we employ anonymous annotation data to learn a model to cluster human experts, and use the clustering results for learning to defer to clusters.

Given the differences, crowdsource or consensus label methods are incompatible with the setting we considered in this paper. In addition, modification of these methods to adapt to the setting of anonymously annotated data is non-trivial. Due to this reason, we do not include these methods as baselines in our evaluation presented in Sec. 4.