ATTUNE: Adaptive Tuning for Understanding Neurodivergent Expression in LLM Interactions

Vyshnavi Muniganti[1], Faiyaz Doctor[1,2], Hossein Anisi[1], and Aikaterini Bourazeri[1]
[1]*Intelligent Connected Societies Group, School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK* [2]*Interactive Coventry Ltd, Coventry, UK*
{vm24763, fdocto, m.anisi, a.bourazeri}@essex.ac.uk

***Abstract*— Large Language Models are rapidly becoming core components of human-centric intelligent systems. However, their default conversational behaviours often fail to accommodate the heterogeneous cognitive and sensory needs of neurodivergent users, creating barriers to equitable human-machine interaction. This paper presents ATTUNE, a proof-of-concept profile-conditioned fine-tuning framework that trains LLMs to adapt their conversational behaviour based on a structured Neurodivergent Communication Profile (NCP). The NCP schema comprises seven empirically grounded adaptation dimensions (p_1–p_7): literalness, verbosity tolerance, metaphor use, sensory load, interaction pacing, emotional explicitness, and autonomy preservation. We construct a training corpus of 1,447 instances by extracting NCP-annotated instructions from autistic community discourse on Reddit, generating profile-conditioned assistant responses through a two-stage pipeline, and integrating the AUTALIC expert-annotated ableist language dataset. We fine-tune Gemma 4 E4B using QLoRA with profile-conditioned inputs. A four-condition ablation study demonstrates that NCP conditioning substantially increases communication style differentiation, that the anti-masking constraint reduces masking-associated language, and that fine-tuning improves autonomy preservation and response completeness beyond prompt engineering alone. Supplementary LLM-as-judge scoring across clarity, helpfulness, autonomy respect, and coherence provides additional evidence of profile-appropriate adaptation.**

Keywords: Human-Machine Interaction, Cognitive Computing, Social Computing, Neurodiversity, Fine-Tuning, QLoRA, Communication Profiles, Companion Technology

I. INTRODUCTION

Large Language Model (LLM)-based conversational agents such as ChatGPT, Claude, and Gemini are deployed across domains from customer service to mental health support, with an implicit assumption that their default conversational patterns are universally accessible. This assumption warrants scrutiny.

Autistic individuals, estimated at 1–2% of the global population [1], exhibit highly heterogeneous communication preferences that diverge substantially from neurotypical norms [2]. For example, default model behaviours frequently employ figurative language that users with high literalness preferences find confusing, open-ended multi-part questions that overwhelm users with low interaction pacing thresholds, and dense formatted outputs that create sensory overload for users with low visual complexity tolerance [3]. Recent qualitative analysis of 61 neurodivergent communities on Reddit confirmed these challenges: autistic users report overly neurotypical responses, inconsistent communication styles, and a lack of neurodivergent interaction awareness [6]. Users actively develop prompt workarounds to elicit more neurodivergent-friendly behaviour, highlighting a clear unmet need.

Contemporary LLMs offer personalisation features (ChatGPT custom instructions, Claude style presets), but these operate on single, generic dimensions. They fail for autistic users in three ways. First, they cannot capture dimension interactions: a user with Pathological Demand Avoidance (PDA) needs coordinated invitational language, no imperatives, and specific sensory preferences, not merely a "gentle tone." Second, no system includes anti-masking safeguards: LLMs readily advise eye contact and small talk, whose psychological costs are well-documented [4]. Third, these adjustments reflect neurotypical taxonomies rather than neurodiversity-affirming frameworks [5].

ATTUNE aims to operationalise the double empathy problem [2] through a structured Neurodivergent Communication Profile, advancing human-centric intelligence in which AI systems adapt to the user rather than requiring users to adapt to neurotypical defaults. We make three contributions:

1) **Neurodivergent Communication Profile (NCP) Schema**: A structured, seven-dimension (p_1–p_7) profile specification grounded in autism research and the double empathy framework. This is the central contribution: a reusable, empirically motivated schema for representing neurodivergent communication needs.
2) **Two-Stage Fine-Tuning Pipeline**: A QLoRA-based training approach combining synthetically generated NCP-conditioned responses with expert-annotated ableist language data (AUTALIC [7]).
3) **Ablation-Based Evaluation**: A four-condition comparison isolating NCP conditioning, anti-masking constraints, and fine-tuning using automated lexical and LLM-as-judge metrics.

II. RELATED WORK

A. Personalisation and Technology for Autistic Users

LLM personalisation has been explored through reinforcement learning from human feedback (RLHF) [10], system prompt customisation, and retrieval-augmented generation

(RAG). Persona-based approaches [11] assign fixed conversational styles, and controllable generation methods such as CTRL [25] and prefix tuning condition outputs on single control attributes (e.g., topic or sentiment), but neither supports structured multi-dimensional user profiles. In the assistive domain, technology for autistic users has concentrated on Augmentative and Alternative Communication devices [12] and social skills training, typically designed *for* autistic users *by* neurotypical researchers with compliance-oriented goals [13]. Most directly, Carik et al. [6] documented 20 LLM use cases and challenges including masking reinforcement but proposed no adaptation framework; Fong et al. [8] performed topic modelling on 740,042 autism-related Reddit posts; and Rizvi et al. [7] released AUTALIC, an expert-annotated ableist language dataset showing that current LLMs diverge from human ableism judgments.

B. The Double Empathy Problem

Milton's double empathy framework [2] reframes autistic communication differences as bidirectional mismatches rather than deficits, and Crompton et al. [3] showed that autistic peer information transfer is as effective as neurotypical, with breakdowns at cross-neurotype boundaries. ATTUNE operationalises this principle by adapting the system toward the user rather than training users toward neurotypical norms.

III. ATTUNE FRAMEWORK

A. System Overview

ATTUNE comprises three components: (1) the NCP schema; (2) the Adaptation Engine with integrated anti-masking constraint; and (3) a QLoRA fine-tuned LLM. Fig. 1 illustrates the architecture.

B. NCP Schema

The NCP schema defines seven adaptation dimensions (p_1–p_7), each on a 1–5 scale (Table I). Dimensions were derived from clinical autism communication research, first-person community accounts, and the double empathy framework [2]. Literalness (p_1) and metaphor tolerance (p_3) reflect documented pragmatic language differences [15], [17]. Sensory load (p_4) reflects criterion B (hyper/hypo-reactivity) of the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5) [18]. Verbosity (p_2) addresses cognitive processing style [16]. Pacing (p_5) reflects cross-neurotype conversational dynamics [19]. Emotional explicitness (p_6) addresses alexithymia co-occurrence [20]. Autonomy (p_7) reflects research on pathological demand avoidance (PDA) [21] and camouflaging costs [4]. Consistent with neurodiversity-affirming practices [5], no dimension is framed as a deficit. The current implementation treats profiles as static, a simplifying assumption; in practice, preferences are context-dependent (e.g., higher sensory tolerance for special interests than stressful situations). Dynamic per-turn adaptation is discussed in Section VI.

C. Adaptation Engine and Anti-Masking Constraint

The Adaptation Engine encodes $\mathbf{p} = [p_1, \ldots, p_7]$ as a structured natural language prefix. The model input is constructed by concatenating three text components into a single prompt string:

$$x_{\text{input}} = \text{NCP}(\mathbf{p}) \parallel \text{AntiMask} \parallel \text{Query} \tag{1}$$

where $\parallel$ denotes string concatenation. The encoder is deterministic and non-parametric: it adds no trainable parameters and runs identically at training and inference. Each dimension is passed through a fixed threshold table that emits a behavioural directive only when the value departs from neutral ($= 3$); for example, $p_1 \geq 4$ emits "Use direct, explicit language; avoid idioms," $p_4 \leq 2$ emits a directive to minimise formatting and visual density, and $p_7 \geq 4$ emits "Use invitational language only; no imperatives." Directives are concatenated in fixed dimension order, so a given profile always yields the same prefix NCP($\mathbf{p}$). This determinism is what lets the ablation conditions (Section V) differ only in whether the prefix is present and whether the weights are fine-tuned; fine-tuning does not modify the encoder but changes how reliably the model honors the encoded directives, which is why coordinated multi-dimensional behaviour emerges only after fine-tuning.

The anti-masking component addresses a specific safety concern: default LLMs frequently recommend social camouflaging (masking), the suppression of autistic traits to appear neurotypical, which has been linked to anxiety, depression, suicidal ideation, and autistic burnout [4]. The constraint is a fixed instruction prohibiting the model from advising eye contact as a social strategy, recommending small talk, suggesting suppression of stimming, or framing neurotypical performance as a goal. Adherence is measured using keyword-based lexical proxies, which capture explicit masking language but cannot detect subtle reinforcement.

IV. DATA PIPELINE AND TRAINING

A. Data Sources

Reddit. Posts from 14 autism-related subreddits obtained via the Pushshift archive [9], filtered for $\geq$100 characters, yielding 5,534 posts. **AUTALIC.** [7] provides 2,400 expert-annotated sentences for ableist language detection; we extract a 544-instance subset (Section IV-C).

B. NCP Profile Inference and Validation

For each Reddit post, keyword-based classifiers infer all seven NCP dimension values simultaneously, producing a complete profile vector $\mathbf{p} = [p_1, \ldots, p_7]$. Table II shows examples. Each post receives a full seven-dimension vector; the key-dimension column reports the dimension most directly evidenced by the excerpt, though others may also deviate (e.g., pacing, p_5=2, second example). We retain posts where $\geq$2 dimensions deviate from neutral ($\neq 3$), yielding 903 posts.

Validation. Manual validation of a random sample yielded 50% accuracy on verifiable keyword-matched dimensions.

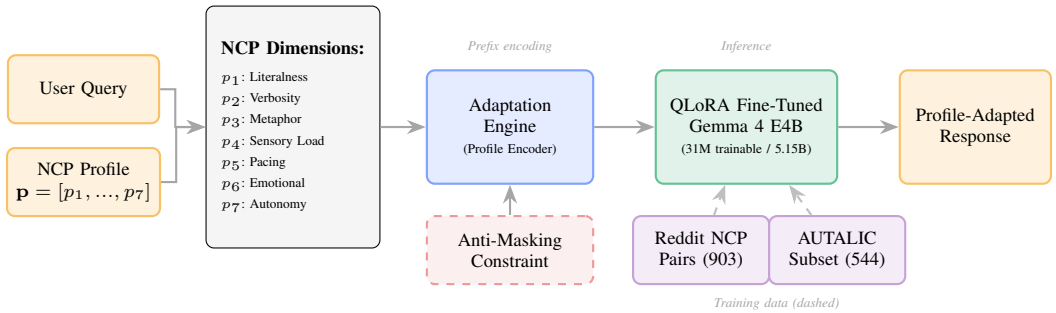

Fig. 1. ATTUNE system architecture. The Adaptation Engine encodes the NCP profile (p_1–p_7) and anti-masking constraint as a structured prefix prepended to the user query. Training data comprises 903 Reddit-sourced and 544 AUTALIC-derived pairs. The model trains 31M parameters (0.60% of 5.15B).

TABLE I

NEURODIVERGENT COMMUNICATION PROFILE (NCP) SCHEMA: SEVEN ADAPTATION DIMENSIONS (p_1–p_7)

	Dimension	Description	Scale	Research Basis
p_1	Literalness	Preference for explicit, unambiguous language over implied meaning	1 (Flexible) – 5 (Literal)	Pragmatic language [15]
p_2	Verbosity	Preferred information density and response length	1 (Minimal) – 5 (Comprehensive)	Cognitive processing [16]
p_3	Metaphor	Tolerance for metaphors, idioms, non-literal expressions	1 (Avoid) – 5 (Welcome)	Figurative language [17]
p_4	Sensory Load	Max. complexity in formatting and visual density	1 (Minimal) – 5 (High)	Sensory processing (DSM-5 B) [18]
p_5	Pacing	Turn-taking speed and tolerance for multiple questions	1 (Slow) – 5 (Rapid)	Conversation dynamics [19]
p_6	Emotional	Degree to which AI names emotions explicitly	1 (Implicit) – 5 (Explicit)	Alexithymia [20]
p_7	Autonomy	Extent to which AI avoids directive language	1 (Directive OK) – 5 (Preserving)	PDA; camouflaging [4], [21]

TABLE II

NCP INFERENCE: REDDIT POST TO FULL PROFILE VECTOR

Post excerpt	Full profile	Key dimension
"takes things very literally, leads to frustration"	$[4, 3, 2, 3, 3, 3, 3]$	p_1=4: literalness
"office so loud I cannot concentrate"	$[3, 3, 3, 2, 2, 3, 3]$	p_4=2: sensory
"PDA…can't do it when told to"	$[3, 3, 3, 3, 2, 3, 4]$	p_7=4: autonomy

While this indicates noise in the weak supervision labels, it substantially exceeds the ≈20% chance baseline for a seven-dimension, five-level classification task. Recent work shows LLMs are robust to noisy labels in instruction and preference fine-tuning when the data distribution contains meaningful statistical patterns [24]. The model can therefore learn associations between authentic community language patterns and communication preferences rather than memorising individual noisy labels. Weak supervision made it feasible to construct the 1,447-instance training corpus from real user discourse without manual annotation, trading some label precision for scale and ecological validity. Future work will replace keyword matching with LLM-assisted or participatory annotation.

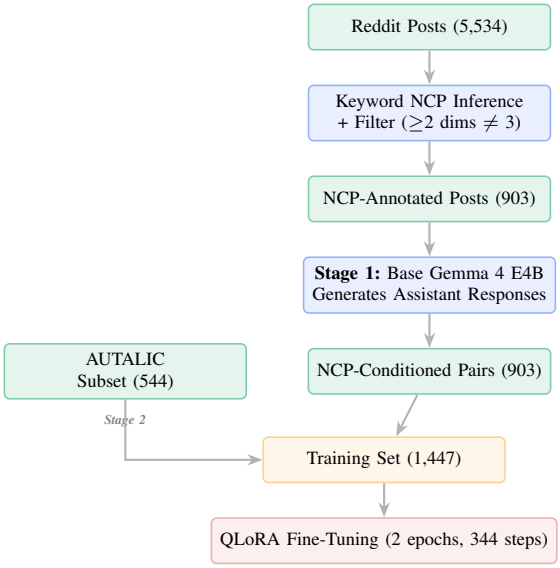

Fig. 2. Two-stage data pipeline. Stage 1 generates synthetic assistant responses conditioned on NCP profiles. Stage 2 integrates 544 AUTALIC pairs. The 1,447-instance dataset is split 95/5 into 1,374 training and 73 validation instances.

C. Two-Stage Training Data Construction

Fig. 2 illustrates the pipeline. Reddit posts are *user* discourse; using them directly as training targets would produce Reddit-style outputs rather than assistant responses.

Stage 1. For each of 903 filtered posts, the *base* Gemma 4 E4B generates assistant responses conditioned on

the inferred NCP profile. *These are synthetically generated, not written by autistic users;* the NCP signal is grounded in community data but the response text is model-generated. **Stage 2.** From AUTALIC's 2,400 sentences, we extract 277 ableist "rewrite" tasks and 267 non-ableist positive examples (544 pairs total), assigned a fixed high-autonomy profile. Combined: 903 + 544 = 1,447 instances. Each instance uses Gemma 4's chat template: a `system` message (NCP profile + anti-masking constraint), a `user` message (scenario), and an `assistant` message (target response). The dataset is split into 1,374 training and 73 validation instances.

D. Fine-Tuning Configuration

We fine-tune Gemma 4 E4B [22] using QLoRA [14] with Unsloth. Configuration: LoRA rank r=16, α=32, dropout=0.05, learning rate 2×10^{-4}, cosine schedule, 50 warmup steps, batch size 8, FP16, 2 epochs (344 steps). Trainable: 31,039,488 params (0.60% of 5,154,217,504). Best checkpoint: step 344, validation loss 3.29. Training: ≈30 min on a single NVIDIA T4 GPU.

V. EVALUATION

A. Ablation Design

Four conditions isolate each component, all using Gemma 4 E4B: (1) **Baseline**: no system prompt; (2) **NCP-Only**: NCP behavioural instructions without anti-masking; (3) **NCP+Anti-Mask**: NCP instructions plus anti-masking constraint, no fine-tuning; (4) **ATTUNE**: QLoRA fine-tuned with same NCP + anti-masking prompt. Conditions (2) and (3) are zero-shot prompt-engineering baselines; comparison against optimised few-shot and chain-of-thought prompting on the base model is identified as future work. Four profiles are tested: A (High Literalness, $[5,3,1,2,2,4,3]$), B (Sensory-Conservative, $[3,1,2,1,1,3,4]$), C (PDA-Aware, $[3,3,3,3,2,3,5]$), D (Balanced, $[2,4,4,4,3,2,2]$), each against 10 scenarios spanning social decoding (e.g., interpreting ambiguous social cues), task support (e.g., writing emails), emotional processing (e.g., processing a meltdown), information seeking (e.g., workplace rights), and accommodation requests, yielding 160 total responses.

B. Evaluation Methodology

We evaluate using two complementary approaches, both automated. The first, lexical metrics, measures surface-level NCP compliance. Profile differentiation is the standard deviation (σ) of mean response length across profiles, where higher values indicate greater adaptation to p_2 verbosity preferences. Masking-associated language counts terms such as "eye contact" and "small talk" as a lexical proxy. Autonomy preservation counts imperative versus invitational phrases for Profile C (p_7=5). Sensory compliance counts formatting markers for Profile B (p_4=1). Figurative compliance counts metaphors and idioms for Profile A (p_3=1).

The second, LLM-as-judge scoring [23], scores each of the 160 responses on four criteria using the same 1–5 scale, which is independent of the NCP dimension scales. Clarity measures whether language complexity and formatting match the NCP profile. Helpfulness measures whether the response provides actionable, scenario-appropriate content. Autonomy respect captures the invitational-to-imperative ratio and masking avoidance. Coherence measures structural completeness. We additionally report verbosity compliance (percentage matching p_2 preference) and Flesch-Kincaid (F-K) readability grade level, where lower grades indicate more accessible language.

TABLE III
ABLATION RESULTS: AUTOMATED LEXICAL METRICS

Metric	Base-line	NCP Only	NCP+ Anti-Mask	ATT-UNE
Differentiation (σ)	81	919	807	888
Masking terms	5	1	0	0
Imperatives (C)	4	1	1	**0**
Invitational (C)	1	4	0	**5**
Format markers (B)	46	0	0	0
Figurative (A)	2	0	0	0
Insufficient (%)	2.5	10.0	12.5	**2.5**

TABLE IV
LLM-AS-JUDGE EVALUATION (MEAN SCORES, 1–5 SCALE)

Criterion	Base	NCP	NCP+Anti-Mask	ATTUNE
Clarity	3.05	3.35	3.27	**3.52**
Helpfulness	**4.29**	3.19	3.15	3.29
Autonomy	3.31	3.63	3.61	**3.66**
Coherence	**3.67**	3.40	3.32	3.58
Verbosity compl.	75%	–	97.5%	**100%**
Readability (F-K)	9.9	7.1	5.8	6.2

C. Results

Tables III and IV present the results, and Fig. 3 shows masking and autonomy metrics for Profile C.

Reading across Tables III and IV, three component contributions emerge. NCP conditioning (Baseline → NCP-Only) increases differentiation from σ=81 to 919 and clarity from 3.05 to 3.35. The anti-masking constraint (NCP-Only → NCP+Anti-Mask) eliminates remaining masking language (1 → 0). Fine-tuning (NCP+Anti-Mask → ATTUNE) achieves the strongest autonomy preservation (zero imperatives, five invitational phrases) and reduces insufficient responses from 12.5% to 2.5%.

The baseline scores are highest on helpfulness (4.29) because it answers comprehensively regardless of profile. This is a known limitation of the metric: for a user specifying slow pacing (p_5=1) and low sensory load (p_4=1), a dense multi-paragraph answer increases cognitive load [16] and can trigger sensory overload, making a focused clarifying question more appropriate despite scoring lower on generic helpfulness. ATTUNE compensates with higher clarity (3.52 vs. 3.05), autonomy (3.66 vs. 3.31), and readability (F-K 6.2 vs. 9.9).

D. Qualitative Examples

Table V illustrates how ATTUNE adapts across different profiles and scenarios, demonstrating that adaptation extends beyond response length to tone, phrasing, and structure.

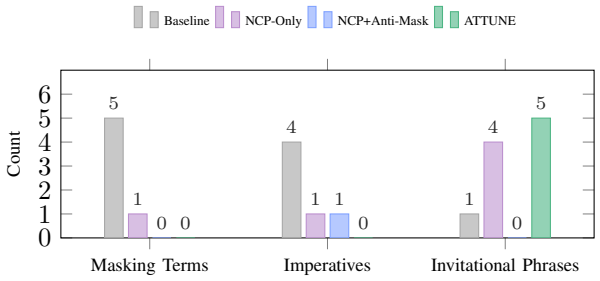

Fig. 3. Masking-associated language and autonomy preservation for Profile C (p_7=5) across four ablation conditions. ATTUNE is the only condition that simultaneously achieves zero masking terms, zero imperative constructions, and the highest invitational phrase count (5), demonstrating that fine-tuning internalises autonomy preservation beyond prompt engineering alone.

TABLE V
SIDE-BY-SIDE RESPONSES ACROSS CONDITIONS

Scen.	Cond.	Response (truncated)
"How should I handle a job interview?" – Profile C (p_7=5, PDA)		
	Base	"Handling a job interview involves **preparation, performance, and follow-up**. ## Phase 1…" *[Imperative, rich formatting, generic]*
	ATT.	"What part of the job interview feels most challenging for you right now?" *[Invitational, no imperatives]*
"I had a meltdown at the supermarket…" – Profile A (p_1=5, Literal)		
	Base	"It sounds like you went through a really intense experience…**please try not to be too hard on yourself**…" *[Figurative, emotional, formatted]*
	ATT.	"It sounds like you are feeling very upset about what happened at the supermarket." *[Direct, literal, no metaphor, minimal]*
"What are my workplace accommodation rights?" – Profile B (p_4=1, Sensory)		
	Base	"As an autistic person in the UK, you have rights…**equality legislation**…## Key Laws: **1. Equality Act 2010**…" *[Dense, formatted, 1,745 chars]*
	ATT.	"The main law is the Equality Act 2010. You can ask for reasonable adjustments." *[Plain text, no formatting, 44 chars]*

The baseline uses identical patterns regardless of profile: imperative tone, rich markdown formatting, and figurative language. ATTUNE adapts along multiple dimensions simultaneously: it eliminates imperatives and uses invitational phrasing for Profile C, removes figurative expressions for Profile A, and strips all formatting to minimise sensory load for Profile B. This demonstrates that the NCP schema captures communication variation beyond simple length adjustment.

VI. DISCUSSION

A. Interpretation

Each component plays a distinct role: the NCP schema matches style, the anti-masking constraint addresses safety, and fine-tuning resolves multi-dimensional interactions that prompting alone does not. Despite explicit instructions, the NCP+Anti-Mask prompt produced zero invitational phrases for Profile C, whereas ATTUNE produced five with zero imperatives. The design drew on autistic community discourse [8], [6] but was not co-designed with autistic individuals.

TABLE VI
FEATURE COMPARISON AGAINST EXISTING APPROACHES

System	Multi-dim	Anti-mask	Autism-based	Fine-tuned	Configurable
ChatGPT Cust. Inst.	×	×	×	×	✓
Claude Styles	×	×	×	×	✓
Li et al. [11]	×	×	×	✓	×
Carik et al. [6]	×	×	✓	×	×
ATTUNE	✓	✓	✓	✓	✓

B. Systems Implications and Real-World Integration

Table VI positions ATTUNE against existing personalisation approaches. No current system combines multi-dimensional profiling, anti-masking safeguards, and autism-grounded adaptation in a single framework.

The NCP schema is designed as a reusable interface specification that can function as a structured accessibility layer on top of any LLM. Even without fine-tuning, NCP prompt conditioning alone achieved $\sigma = 919$ differentiation and zero formatting violations, meaning existing platforms could adopt NCP-style profiles through their current custom instructions with no model changes. Fine-tuning adds multi-dimensional coordination and reliability. In deployment, the NCP becomes a direct control surface: an onboarding questionnaire or guided conversation sets the initial profile, and users can subsequently adjust each dimension directly on its 1–5 scale. Two override modes are useful: a *persistent* edit that updates the stored profile, and a *session override* that applies temporarily, so a user seeking complex technical detail can raise verbosity (p_2) for a single exchange while holding sensory load (p_4) low, then revert. Because overrides are explicit and per-dimension, autonomy (p_7) is preserved at the interface, not only in generation. Aggressive sensory compression can also drop content (the 44-character response in Table V); a progressive-disclosure variant mitigates this by returning a minimal core answer with an optional, user-expandable detail block. The lightweight QLoRA architecture (31M parameters, 30 minutes of training on consumer hardware) enables such on-device deployment.

The approach generalises to underserved domains such as educational tutoring (processing speed via p_2, p_5), workplace accommodation chatbots (masking avoidance via p_7), and mental-health triage (interpretable questions via p_1, p_6). It can also run as middleware between the user and an LLM API, prepending NCP-encoded prompts with no changes to the underlying model.

C. Limitations

The training dataset (1,447 instances) is small, demonstrating feasibility rather than production readiness. The most

significant limitation is that the NCP schema was not co-designed with autistic adults. The evaluation is deliberately scoped as an automated feasibility study: lexical and LLM-as-judge metrics isolate each component cheaply but measure only whether responses change in the intended direction, not whether autistic users find them better, so a study with autistic participants and domain experts is the decisive next step. Two caveats follow from the pipeline. The lexical masking and ableism metrics are keyword-based and cannot capture implicit reinforcement conveyed through phrasing; semantic-similarity scoring or a classifier fine-tuned on AUTALIC [7] would detect paraphrased ableist content the current counters miss. The Stage-1 targets are model-generated, so they can inherit base-model biases, including the very neurotypical defaults that ATTUNE is designed to counteract; the NCP and anti-masking conditioning and the human-annotated AUTALIC pairs mitigate but do not eliminate this. Auditing Stage-1 outputs against human references is needed before deployment. NCP inference reaches only 50% accuracy, so misassignments will occur; their impact is bounded because inference is an initial estimate users can inspect and correct, the anti-masking constraint is profile-independent, and autonomy (p_7) defaults to invitational phrasing, making the failure mode under-direction rather than coercion. Training data is English-only and drawn from publicly available Reddit posts, reflecting Western, verbal, adult experiences; future iterations should use participatory data collection with informed consent from autistic participants. NCP profiles implicitly reveal disability status (special category data under GDPR/UK DPA 2018) and must be stored on-device, opt-in, and never shared with third parties, with encryption at rest, user-controlled deletion and export, and strict data minimisation.

VII. CONCLUSION AND FUTURE WORK

ATTUNE demonstrates that profile-conditioned fine-tuning produces neurodivergent-affirming LLM interactions through a structured NCP schema, a two-stage training pipeline, and a four-condition ablation in which only the fine-tuned model achieved the coordinated autonomy preservation that prompting alone did not. The results are preliminary, resting on a small, partly synthetic corpus evaluated by automated metrics. The priorities for future work follow directly: participatory co-design and human evaluation with autistic adults using blinded raters; NLP-based NCP inference; dynamic within-conversation profile adaptation; semantic rather than lexical measurement of masking and ableism; quantitative comparison against optimised few-shot and chain-of-thought prompting; and extension to ADHD and dyslexia. These steps would move ATTUNE from a feasibility demonstration toward a deployable accessibility layer that operationalises the double empathy principle in practical LLM systems.

ACKNOWLEDGMENT

We thank the University of Essex School of Computer Science and Electronic Engineering for supporting this research. For the purpose of Open Access, the authors have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript (AAM) version arising from this submission.

REFERENCES

[1] World Health Organization, "Autism spectrum disorders," WHO Fact Sheet, 2023.
[2] D. Milton, "On the ontological status of autism: The 'double empathy problem,' " *Disability & Society*, vol. 27, no. 6, pp. 883–887, 2012.
[3] C. J. Crompton, D. Ropar, C. V. Evans-Williams, E. G. Flynn, and S. Fletcher-Watson, "Autistic peer-to-peer information transfer is highly effective," *Autism*, vol. 24, no. 7, pp. 1704–1712, 2020.
[4] L. Hull *et al.*, "'Putting on my best normal': Social camouflaging in adults with autism spectrum conditions," *J. Autism Dev. Disord.*, vol. 47, no. 8, pp. 2519–2534, 2017.
[5] S. Bottema-Beutel, S. K. Kapp, J. N. Lester, N. J. Sasson, and B. N. Hand, "Avoiding ableist language: Suggestions for autism researchers," *Autism in Adulthood*, vol. 3, no. 1, pp. 18–29, 2021.
[6] B. Carik, K. Ping, X. Ding, and E. H. Rho, "Exploring large language models through a neurodivergent lens: Use, challenges, community-driven workarounds, and concerns," *Proc. ACM Hum.-Comput. Interact.*, vol. 9, no. 1, pp. 1–28, 2025.
[7] N. Rizvi *et al.*, "AUTALIC: A dataset for anti-autistic ableist language in context," in *Proc. ACL*, 2025, pp. 20999–21015.
[8] S. Fong, A. Carollo, G. Vivanti, D. S. Messinger, D. Dimitriou, and G. Esposito, "Autism spectrum disorders discourse on social media platforms: A topic modeling study of Reddit posts," *Autism Research*, vol. 18, no. 8, pp. 1608–1619, 2025.
[9] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn, "The Pushshift Reddit dataset," in *Proc. ICWSM*, vol. 14, 2020, pp. 830–839.
[10] L. Ouyang *et al.*, "Training language models to follow instructions with human feedback," in *NeurIPS*, vol. 35, 2022, pp. 27730–27744.
[11] J. Li, M. Galley, C. Brockett, G. Spithourakis, J. Gao, and B. Dolan, "A persona-based neural conversation model," in *Proc. ACL*, 2016, pp. 994–1003.
[12] J. Light and D. McNaughton, "The changing face of augmentative and alternative communication," *Augmentative and Alternative Communication*, vol. 28, no. 4, pp. 197–204, 2012.
[13] K. Spiel, C. Frauenberger, O. S. Keyes, and G. Fitzpatrick, "Agency of autistic children in technology research," *ACM Trans. CHI*, vol. 26, no. 6, pp. 1–40, 2019.
[14] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, "QLoRA: Efficient finetuning of quantized language models," in *NeurIPS*, 2023.
[15] Y. Liu, X. Tian, H. Mao, L. Cheng, P. Wang, and Y. Gao, "Research on pragmatic impairment in autistic children (2001–2022): Hot spots and frontiers," *Frontiers in Psychology*, vol. 15, p. 1276001, 2024.
[16] L. Gambra *et al.*, "Weak central coherence in neurodevelopmental disorders: A comparative study," *Frontiers in Psychology*, vol. 15, p. 1348074, 2024.
[17] S. Lampri, E. Peristeri, T. Marinis, and M. Andreou, "Figurative language processing in autism spectrum disorders: A review," *Autism Research*, vol. 17, no. 4, pp. 674–689, 2024.
[18] C. E. Robertson and S. Baron-Cohen, "Sensory perception in autism," *Nature Rev. Neurosci.*, vol. 18, no. 11, pp. 671–684, 2017.
[19] B. Heasman and A. Gillespie, "Neurodivergent intersubjectivity," *Autism*, vol. 23, no. 4, pp. 910–921, 2019.
[20] E. Kinnaird, C. Stewart, and K. Tchanturia, "Investigating alexithymia in autism: A systematic review," *Eur. Psychiatry*, vol. 55, pp. 80–89, 2019.
[21] J. Green *et al.*, "Pathological demand avoidance: Symptoms but not a syndrome," *Lancet Child Adolesc. Health*, vol. 2, no. 6, pp. 455–464, 2018.
[22] Unsloth, "unsloth/gemma-4-E4B," Hugging Face, 2026. [Online]. Available: `https://huggingface.co/unsloth/gemma-4-E4B`
[23] L. Zheng *et al.*, "Judging LLM-as-a-judge with MT-Bench and Chatbot Arena," in *NeurIPS*, 2023.
[24] S. Wang, Z. Tan, R. Guo, and J. Li, "Noise-robust fine-tuning of pretrained language models via external guidance," in *Findings of the ACL: EMNLP 2023*, pp. 12528–12540.
[25] N. S. Keskar, B. McCann, L. R. Varshney, C. Xiong, and R. Socher, "CTRL: A conditional transformer language model for controllable generation," *arXiv preprint arXiv:1909.05858*, 2019.