

Group	Category English	Description English	Example English	Category German	Description German	Example German
Survey-related	Anonymity	Respondents participate in the survey because of the promised anonymity.	everything remains anonymous	Anonymität	Die Befragten nehmen an der Umfrage teil, weil Anonymität versprochen wurde.	alles bleibt anonym
Survey-related	Brevity	Respondents participate in particular because of the brevity of the questionnaire.	the surveys don't take that long	Kürze	Die Befragten nehmen an der Umfrage teil, weil der Fragebogen kurz ist.	die Umfragen dauern nicht so lange
Intrinsic	Curiosity	Respondents participate in the survey because they are curious about, for example, the current research, the results, or the topics of the survey.	Curiosity	Neugierde	Die Befragten nehmen an der Umfrage teil, weil sie neugierig auf beispielsweise die Forschung, Ergebnisse, oder Themen der Umfrage sind.	Neugier
Extrinsic	Dutifulness	Respondents participate in the survey because they feel morally obligated to do so.	civic sense of duty	Moralische Verpflichtung	Die Befragten nehmen an der Umfrage teil, weil sie sich moralisch dazu verpflichtet fühlen.	bürgerschaftliches Pflichtbewußtsein
Intrinsic	Fun	Respondents participate in the survey because they enjoy it or because they feel some pleasure and comfort from participating.	fun answering the question	Spaß	Die Befragten nehmen an der Umfrage teil, weil es ihnen Spaß macht oder sie es angenehm und Gefallen daran finden.	Spaß am Beantworten der Frage
Intrinsic	Help politicians	Respondents participate in the survey because they want to help politicians or to emphasize the importance and significance of participation in surveys for policymaking	would like to help political leaders	Politik helfen	Die Befragten nehmen an der Umfrage teil, weil sie Politikern helfen wollen, oder betonen die Wichtigkeit der Teilnahme für die Politik.	möchte Verantwortlichen aus der Politik helfen
Intrinsic	Help science	Respondents participate in the survey because they want to help science or emphasize the importance and significance of participation for science.	for social science research	Wissenschaft helfen	Die Befragten nehmen an der Umfrage teil, weil sie der Wissenschaft helfen wollen, oder betonen die Wichtigkeit der Teilnahme für die Wissenschaft.	für die sozialwissenschaftliche Forschung
Intrinsic	Help society	Respondents participate in the survey because they want to help the society or underline the importance of survey participation for society (or economy).	important for society	Gesellschaft helfen	Die Befragten nehmen an der Umfrage teil, weil sie der Gesellschaft helfen wollen, oder betonen die Wichtigkeit der Teilnahme für die Gesellschaft (oder Wirtschaft).	wichtig für die Gesellschaft
Intrinsic	Help, not further specified	Respondents participate in the survey because they want to help but do not further specify to whom this help is directed.	to do others a favor	Helfen, unspezifisch	Die Befragten nehmen an der Umfrage teil, weil sie helfen wollen, aber machen keine Angaben dazu, wem genau geholfen werden soll.	um anderen einen Gefallen zu tun
Other	Importance in general	Respondents talk about importance of participating more broadly.	important	Allgemeinen	Die Befragten nehmen an der Umfrage teil, weil sie es allgemein für wichtig erachten.	wichtig
Extrinsic	Incentive	Respondents participate in the survey because it is rewarded with money.	remuneration	Incentive	Die Befragten nehmen an der Umfrage teil, weil die Teilnahme mit Geld belohnt wird.	Vergütung
Intrinsic	Influence	Respondents have a clear desire to influence policy or research.	To help shape the future	Einfluss nehmen	Die Befragten haben einen eindeutigen Wunsch, Einfluss auf Politik oder Forschung zu nehmen.	Zukunft mitzugestalten
Intrinsic	Interest	Respondents indicate that they are personally interested in the survey's topic, results, or other characteristics.	I find the survey interesting	Interesse	Die Befragten nehmen an der Umfrage teil, weil sie am Thema, den Ergebnissen, oder anderen Eigenschaften der Umfrage interessiert sind.	Ich finde die Umfrage interessant
Intrinsic	Learning	Respondents take part in the survey because they want to educate themselves.	input for thoughts	Weiterbildung	Die Befragten nehmen an der Umfrage teil, weil sie sich weiterbilden wollen.	Gedankeninput
No reason	No reason	Respondents explicitly say that they have no reason to participate.	no important reason	Kein Grund	Die Befragten sagen ausdrücklich, dass sie keinen besonderen Grund haben, warum sie an der Umfrage teilnehmen.	kein wichtiger Grund
Other	Non-identifiable/Other	Answers that cannot be interpreted or put into one of the other categories are classified in this class.	questions	Anderes	Antworten, die nicht in eine der anderen Kategorien passen oder nicht interpretiert werden können.	Fragen
Survey-related	Other survey characteristics	Respondents participate because of the characteristics of the panel. Interest and curiosity regarding the questions (e.g., interesting questions) are not put into this class but in the interest class.	good questions	Anderes, umfragebezogen	Die Befragten nehmen aufgrund von anderen Eigenschaften der Umfrage an ihr teil. Interesse und Neugierde hinsichtlich der darin gestellten Fragen zählen nicht dazu, sondern gehören zur entsprechenden Kategorie (Interesse bzw. Neugierde).	gute Fragen
Survey-related	Professionalism	Respondents participate since they trust the GESIS Institute and think it is a respectable institution.	respectable institute	Seriosität	Die Befragten nehmen an der Umfrage teil, weil sie dem GESIS-Institut vertrauen oder denken, dass es eine angesehene Einrichtung ist.	seriöses Institut
Survey-related	Recruiter	Respondents participate in the interview because of the characteristics of the recruiter or the recruitment interview. In addition to the recruitment, there is a reference to the recruiter or the recruitment interview in this class.	The personal interview was very interesting	Rekrutierer	Die Befragten nehmen aufgrund von Eigenschaften des Rekrutierers oder der Rekrutierungs-Befragung an der Umfrage teil. Zusätzlich zur Rekrutierung an sich erwähnen die Befragten explizit den Rekrutierer oder die Rekrutierungs-Befragung.	Das persönliche vorgrspräch war sehr interessant
Survey-related	Recruitment	Respondents participate in the survey because they were sampled and asked to participate.	I was asked	Rekrutierung	Die Befragten nehmen an der Umfrage teil, weil sie zufällig ausgewählt und gefragt wurden, teilzunehmen.	man hat mich gefragt
Intrinsic	Routine	Respondents participate in the survey because it is a regular occurrence. They started it and want to remain consistent in their behavior.	Regularity	Routine	Die Befragten nehmen an der Umfrage teil, weil sie regelmäßig stattfindet. Sie haben einmal damit begonnen und wollen in ihrem Verhalten konsistent bleiben.	Regelmäßigkeit
Intrinsic	Tell opinion	Respondents take part in the survey because they want to share their opinion.	express my opinion	Meinung mitteilen	Die Befragten nehmen an der Umfrage teil, weil sie ihre Meinung mitteilen wollen.	Meinungsäußerung

Category English	Category German	Example 1 German	Example 1 English	Example 2 German	Example 2 English	Example 3 German	Example 3 English	Example 4 German	Example 4 English
Anonymity	Anonymität	Annonymitaet Habe das bisschen Zeit übrig	anonymity Have that little bit of time to spare	anonym	anonymous	anonymisierte Datenerhebung	anonymized data collection	alles bleibt anonym die Umfragen dauern nicht so lange	anonymous the surveys don't take that long
Brevity	Kürze			geht zügig Gespannt, was es so für Umfragen gibt Ich fühle mich	goes quickly Curious to see what kind of surveys there is	NICH LANG ich bin jedesmal neugierig, was es bürgerschaftliches	NOT LONG I'm curious every time to see what kind of		
Curiosity	Neugierde Moralische	Neugier	Curiosity	verpflichtet	I felt obligated	Pflichtbewußtsein Ich mache gerne Umfragen	civic sense of duty	Bürgerpflicht Spaß am Beantworten der Frage	civic obligation fun answering the question
Dutifulness	Verpflichtung	Verpflichtung	Obligation	es macht spass weil es hilfreich für die Bundesregierung ist	it's fun because it is helpful to the federal				
Fun	Spaß	Spaß möchte	Fun would like to help			wichtig für die Wissenschaft mein beitrug zur Gesellschaft	I like doing surveys	für die sozialwissenschaftliche Umfragen sind für Gesellschaft wichtig	for social science research surveys are important for society
Help politicians	Politik helfen	Wissenschaft Verantwortlichen aus um der Wissenschaft zu helfen	political leaders	Forschung	research		important to science my contribution to society		
Help science	helfen Gesellschaft	wichtig für die Gesellschaft	to help science	für die Gesellschaft um anderen einen Gefallen zu tun	for society	Aus Hilfsbereitschaft Teilnahme ist wichtig für die Politik	out of helpfulness Participation is important for politics	um zu helfen ich finde es wichtig und gut	to help I find it important and good
Help society	Helfen,	möchte helfen	would like to help	Umwelt	environment				
Help, not further specified Importance in general	unspezifisch Bedeutung im Allgemeinen	wichtig	important						
Incentive	Incentive	Belohnung Hoffe, Einfluss auf die Politiker nehmen zu Ich finde die Umfrage	reward Hope to be able to influence politicians I find the survey	gut bezahlt Zur Besserung der jetzigen Lage	well paid To improve the current situation	Prämie von 5EUR Vielleicht was verändern zu können mich interessieren die Fragen	premium of 5EUR Maybe to be able to change something the questions interest me	Vergütung	remuneration To help shape the future the questions and various answers are own evaluation of my behavior
Influence	Einfluss nehmen	interessant selbst Gedanken zu spez. Themen zu	interesting to think for myself on specific topics	aus Interesse	out of interest	Gedankeninput Ich habe keine Gründe	input for thoughts	Zukunft mitzugestalten Die Fragen und verschiedenen Eigene Bewertung meines Verhaltens	
Interest	Interesse			persönliche reflektion Keine Gründe nicht überhaupt nicht	personal reflection no reasons not at all				
Learning	Weiterbildung	kein wichtiger Grund Meine Eltern zwingen mich	no important reason My parents force me	Fragen Gesellschaftsumfrage n weil ich glaube GESIS ist serisös, nehme bin persönlich auf angenehme Art	questions social surveys because I believe GESIS is respectable, was asked personally in a pleasant way	flexibel online ausfüllbar Serioses Auftreten von GESIS Bin angenehm kontaktiert worden Anfrage, ob ich teilnehmen möchte	flexible online completion respectable appearance of GESIS I was contacted in a pleasant way Inquiry if I wanted to participate	Sinnvolle Umfrage, nicht für Werbezwecke	meaningful survey, not for advertising
No reason Non- identifiable/Other Other survey characteristics	Kein Grund Anderes Anderes, umfragebezogen	gute Fragen	good questions respectable interviewer						
Professionalism	Seriosität	Seriöser Fragesteller	interviewer						
Recruiter	Rekrutierer	Bin überredet worden.	Was persuaded						
Recruitment	Rekrutierung	man hat mich gefragt	I was asked	wurde darum gebeten Ich wurde damals gefragt und bin dabei meine Meinung mitteilen	was asked to do this I was asked at the time and stuck with it share my opinion	Fortsetzung des einmal begonnenen kann meine Meinung äußern	Continuation of what was once started can express my opinion	einmal mitgemacht, dann weiter Meine Meinung mit einbringen	participated once, then continued contribute my opinion
Routine	Routine	Regelmäßigkeit	Regularity						
Tell opinion	Meinung mitteilen	Meinungsäußerung	express my opinion						

Appendix II

1. Survey Question

(2) Aus welchen Gründen nehmen Sie an den Umfragen des GESIS GesellschaftsMonitors teil?

Bitte nennen Sie die drei wichtigsten Gründe.

Wichtigster Grund:

Zweitwichtigster Grund:

Drittwichtigster Grund:

Figure A1: Question on survey motivation as asked in the GESIS Panel.pop (Bosnjak et al., 2018, GESIS, 2024).

Translation: "(2) For what reasons do you participate in the surveys of the GESIS GesellschaftsMonitor? Please name the three most important reasons. Most important reason: ... , Second most important reason: ... Third most important reason: ... "

2. Prompting

Du bist eine Expertin für Umfragen, die offene Antworten auf die Frage, wieso Personen an einer Umfrage teilnehmen, klassifiziert. Ordne diese Teilnahmegründe genau einer der folgenden Kategorien zu.

Die Kategorien sind:

*[GERMAN CATEGORY 1: German description]
[GERMAN CATEGORY 2: German description]
[...]
[GERMAN CATEGORY 22: German description]*

Stelle deine bestmögliche Vermutung an, auch wenn es schwer fällt.

Antworte im folgenden Format: Teilnahmegrund | KATEGORIE.

Begründe deine Zuordnung nicht, sondern gib nur den Teilnahmegrund und deine Zuordnung zurück.

Beispiele:

*[German example reason | GERMAN CATEGORY 1]
[German example reason | GERMAN CATEGORY 2]
[...]
[German example reason | GERMAN CATEGORY 22]*

Klassifiziere den folgenden Teilnahmegrund:

[German open-ended response]

Figure A2: German prompt used for LLM-based classifications of the open-ended survey question. Categories and, in the detailed approach, descriptions (green font) were randomized across individual queries. In the few-shot approach, examples (blue font) were randomly selected, the selection being held constant, but presented in random order across queries. For details of descriptions and examples used, see Appendix I.

3. Failed classifications

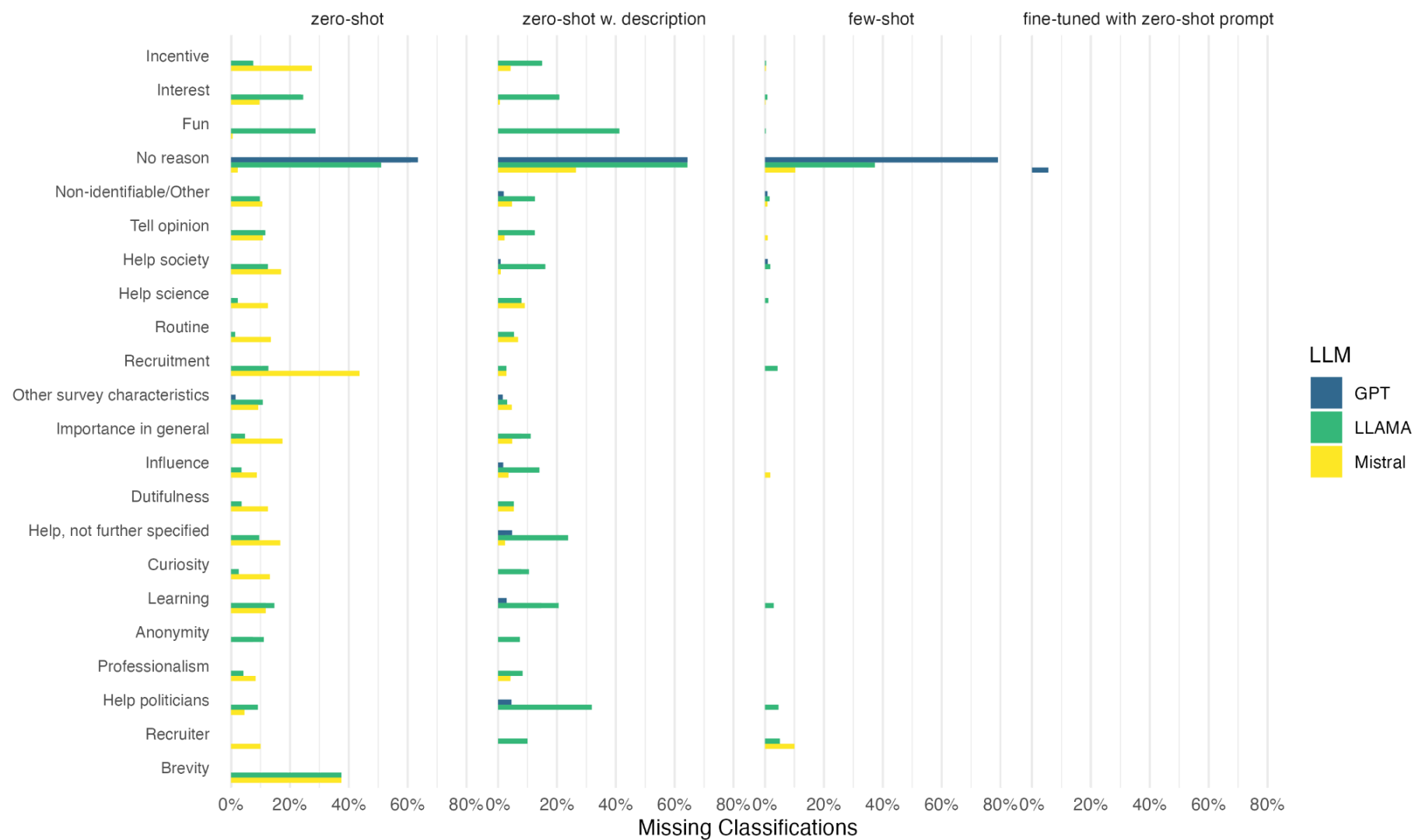


Figure A3: Proportion of missing classifications by category, LLM, and prompting approach.

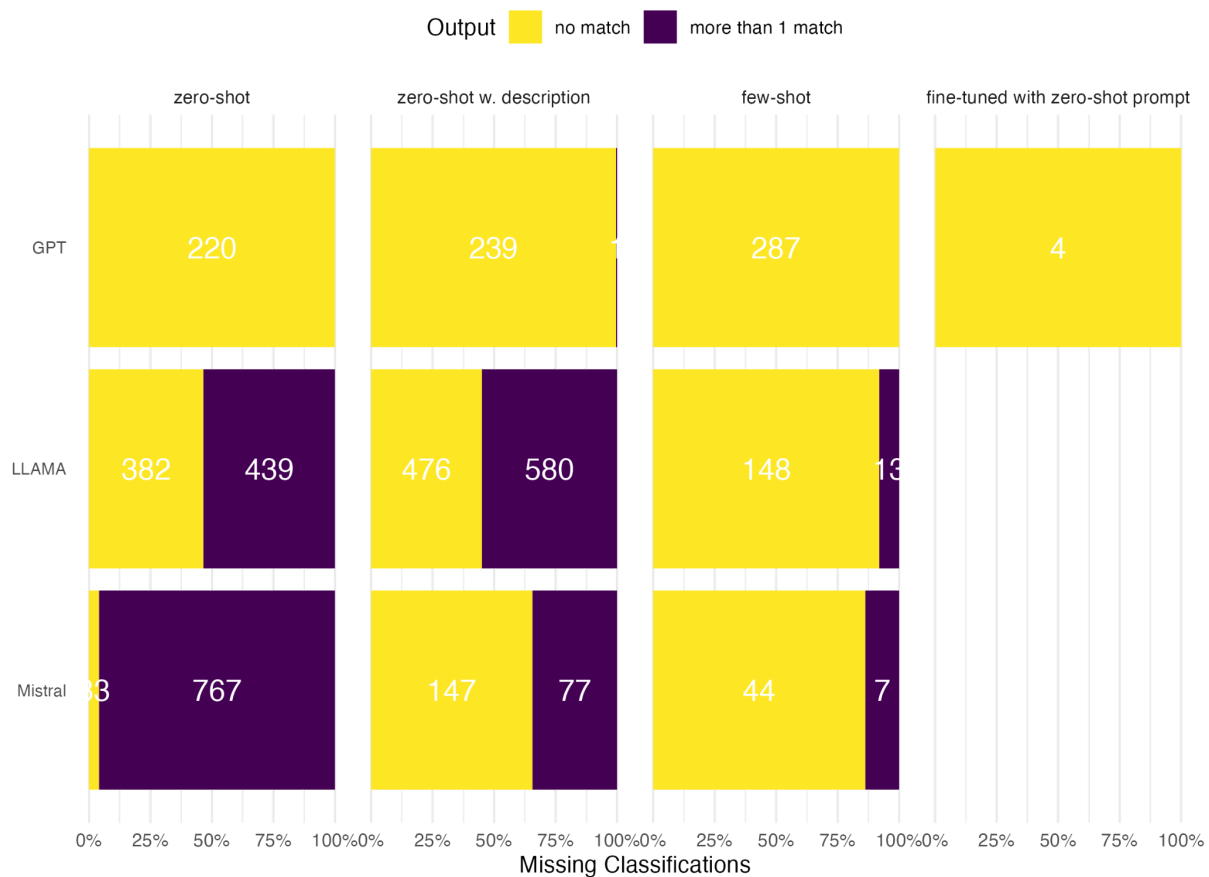


Figure A4: Distribution of outputs recorded as missing due to missing or ambiguous classifications.

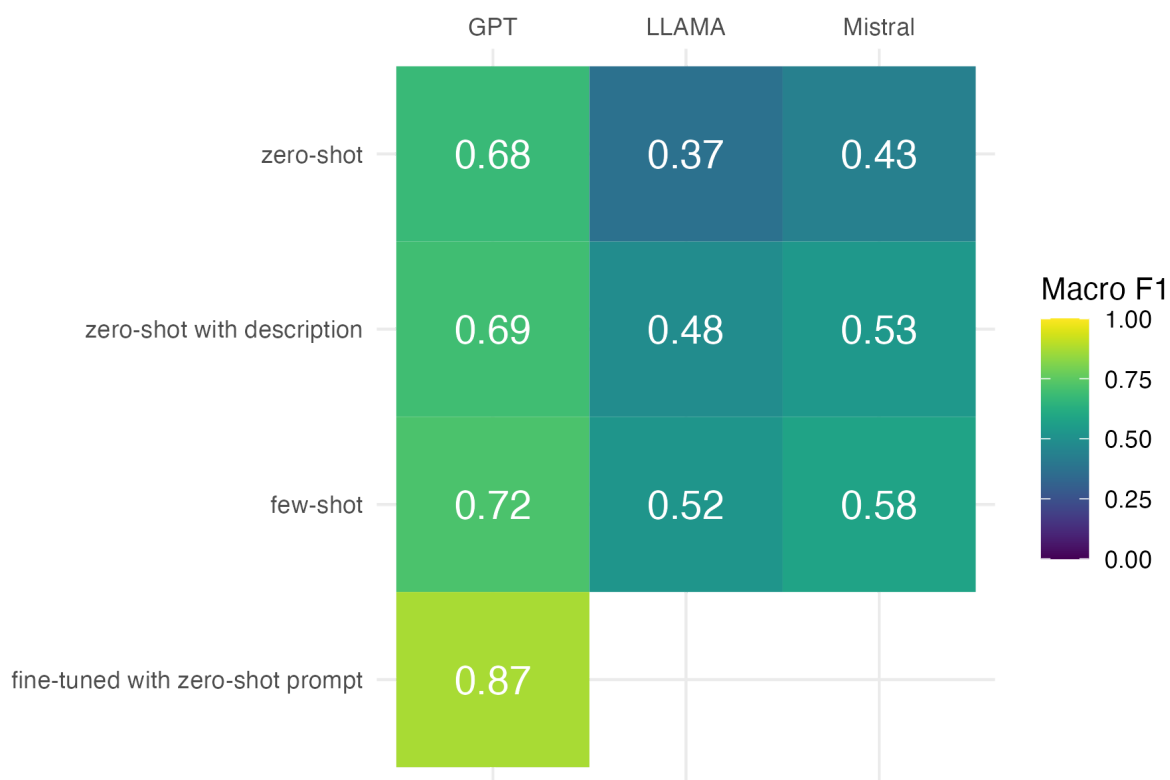


Figure A5: Macro F1 scores by LLM and prompting approach without missing values.

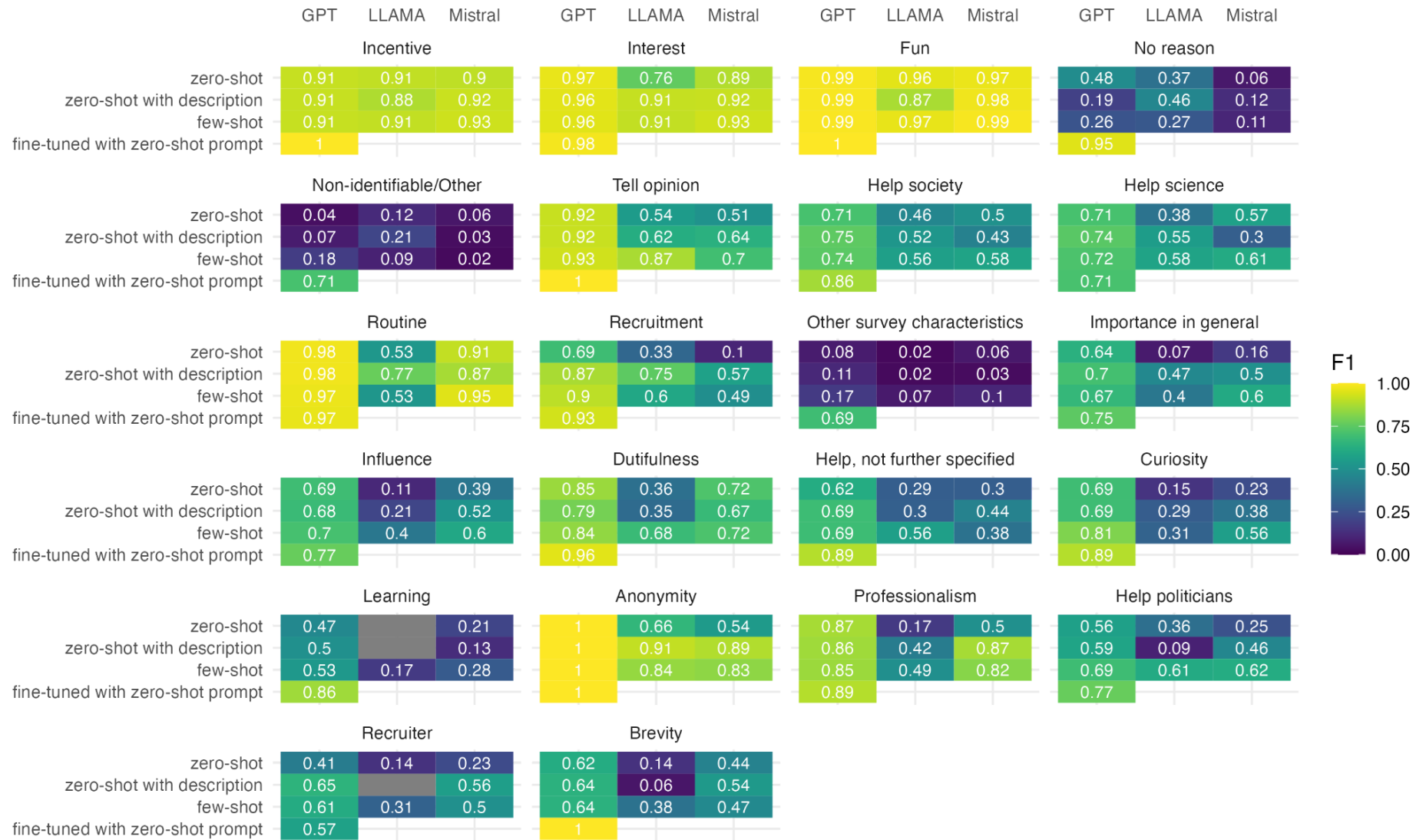


Figure A6: Per-category scores by LLM and prompting approach without missing values.

4. Additional performance metrics

F1 Scores

F1 scores evaluate machine learning model accuracy by considering model precision and recall. Recall refers to the number of correct classifications for a given category (true positives) divided by the number of actual classifications for that category (true positives + false negatives), measuring the share of classifications for a category as coded by the human experts that were correctly classified by the LLM. Precision refers to the number of correct classifications for a given category (true positives) divided by the number of all classifications for that category (true + false positives), measuring the share of LLM classifications for a given category that were correct. The F1 score is defined as $2 * \text{precision} * \text{recall} / (\text{precision} + \text{recall})$, with a range of [0;1].

Macro F1 scores first calculate the F1 for each category separately and then average across these scores for a total model performance score, thereby taking into account unbalanced datasets, which may result in the LLM just assigning the modal category.

Weighted F1 scores adjust for class imbalance by taking a weighted average of per-category F1 scores based on the number of true instances in each category. They are calculated as the weighted harmonic mean of precision and recall. This helps balance the contributions of minority categories without disproportionately emphasizing them. They thus offer a compromise between macro F1 and accuracy (see below), balancing performance on both common and rare classes without over-penalizing a model that performs well on the modal category.

Accuracy

Accuracy describes the proportion of cases correctly classified out of all cases.

Intraclass Correlation Coefficient (ICC)

The ICC quantifies the agreement raters by evaluating the proportion of variance attributable to differences between raters, relative to the total variance. The type used here is ICC(2,1), which assumes a two-way random-effects model and measures agreement. Thus, it does not consider the human-generated classifications as a benchmark or ground truth, but simply as a different coder to compare the LLM to.

Cohen's Kappa

Cohen's Kappa is a measure of agreement that takes into account the agreement occurring by chance, expressed as a proportion of the total possible improvement. It is particularly helpful for imbalanced datasets, as it quantifies model performance relative to a baseline of random chance, i.e., a naive classifier.

Metric	Approach	GPT-4o	Llama-3.2	Mistral NeMo
Weighted F1	<i>zero-shot</i>	0.72	0.40	0.46
	<i>with descriptions</i>	0.72	0.42	0.63
	<i>few-shot</i>	0.72	0.63	0.67
	<i>fine-tuned</i>	0.91		
Accuracy	<i>zero-shot</i>	0.81	0.55	0.60
	<i>with descriptions</i>	0.79	0.58	0.74
	<i>few-shot</i>	0.80	0.74	0.77
	<i>fine-tuned</i>	0.95		
ICC	<i>zero-shot</i>	0.83	0.37	0.61
	<i>with descriptions</i>	0.81	0.47	0.67
	<i>few-shot</i>	0.84	0.65	0.74
	<i>fine-tuned</i>	0.96		
Cohen's Kappa	<i>zero-shot</i>	0.76	0.47	0.53
	<i>with descriptions</i>	0.75	0.50	0.68
	<i>few-shot</i>	0.75	0.68	0.72
	<i>fine-tuned</i>	0.94		

Table A1: Classification performance metrics per LLM and prompting approach, compared to human labels.

5. Confusion Matrices

GPT – zero-shot

R: Prediction / C: Reference	Anonymity	Brevity	Curiosity	Dutifulness	Fun	Help politicians	Help science	Help society	Help, not further specified	Importance in general	Incentive	Influence	Interest	Learning	No reason	Non-identifiable/Other	Other survey characteristics	Professionalism	Recruiter	Recruitment	Routine	Tell opinion
Anonymity	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Brevity	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0
Curiosity	0	0	38	0	0	0	0	0	0	0	0	0	10	3	0	20	1	0	0	0	0	0
Dutifulness	0	0	0	48	0	0	0	1	1	0	0	0	0	0	0	6	0	0	0	0	1	0
Fun	0	0	0	0	376	0	0	0	0	0	0	0	1	1	0	2	5	0	0	0	0	0
Help politicians	0	0	0	0	0	22	1	1	1	0	0	7	2	1	0	21	0	0	0	0	0	0
Help science	0	0	0	0	0	0	79	2	11	12	0	3	1	0	0	26	0	1	0	0	0	0
Help society	0	0	0	0	0	0	0	96	7	5	0	9	1	1	0	40	0	0	0	0	0	0
Help, not further specified	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	2	0	0	0	1	0	0
Importance in general	0	0	0	0	0	0	1	10	0	42	0	0	0	0	0	12	4	0	0	0	0	0
Incentive	0	0	0	5	0	0	0	0	0	0	1523	0	0	0	0	3	0	0	1	0	0	0
Influence	0	0	0	0	0	0	0	0	0	1	0	38	1	0	0	12	0	0	0	0	0	1
Interest	0	0	0	0	0	0	0	1	0	1	0	0	1234	6	0	27	31	0	1	0	0	0
Learning	0	0	0	0	0	0	4	0	1	2	0	0	0	21	0	28	0	0	0	0	0	0
No reason	0	0	0	1	0	0	0	0	0	0	270	0	2	0	134	4	2	0	1	3	0	0

Non-identifiable/Other	0	0	0	1	0	0	0	0	0	0	3	0	0	0	6	5	1	0	1	0	0	0
Other survey characteristics	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	9	3	0	0	0	0	0
Professionalism	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	23	0	0	0	0
Recruiter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	26	0	0
Recruitment	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	4	41	1	0
Routine	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	72	0
Tell opinion	0	0	0	0	0	0	2	1	0	0	0	0	0	1	0	36	1	0	0	0	0	231
NA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	219	0	1	0	0	0	0	0

Table A2: Confusion matrix (actual vs. predicted categories) for GPT under zero-shot prompting.

GPT – with descriptions

R: Prediction / C: Reference	Anonymity	Brevity	Curiosity	Dutifulness	Fun	Help politicians	Help science	Help society	Help, not further specified	Importance in general	Incentive	Influence	Interest	Learning	No reason	Non-identifiable/Other	Other survey characteristics	Professionalism	Recruiter	Recruitment	Routine	Tell opinion
Anonymity	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Brevity	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0
Curiosity	0	0	38	0	0	0	1	0	0	1	0	0	12	1	0	18	1	0	0	0	0	0
Dutifulness	0	0	0	38	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
Fun	0	0	0	0	376	0	0	0	0	0	0	0	0	1	0	2	4	0	0	0	0	0
Help politicians	0	0	0	0	0	17	1	1	0	0	0	2	0	0	0	15	1	0	0	0	0	0
Help science	0	0	0	0	0	0	72	0	5	7	0	2	0	0	0	20	0	0	0	0	0	0
Help society	0	0	0	0	0	1	0	104	7	4	0	8	1	1	0	39	0	0	0	0	0	0
Help, not further specified	0	0	0	0	0	0	0	0	24	0	0	3	0	0	0	3	0	0	0	0	0	0
Importance in general	0	0	0	0	0	0	2	3	0	47	0	1	0	0	0	16	2	0	0	0	0	0
Incentive	0	0	0	6	0	0	0	0	0	0	1522	0	0	0	0	1	0	0	1	0	0	0
Influence	0	0	0	0	0	3	0	2	0	0	0	40	1	0	0	12	0	0	0	0	0	4
Interest	0	0	0	0	0	0	4	1	0	1	0	0	1235	9	0	40	35	0	0	0	0	0
Learning	0	0	0	0	0	0	3	0	0	2	0	0	0	22	0	27	0	0	0	0	0	0
No reason	0	0	0	0	0	0	0	0	0	0	204	0	0	0	36	3	2	0	0	0	0	0

Non-identifiable/Other	0	0	0	1	0	0	0	0	0	0	71	0	2	0	91	16	2	0	0	0	0	0
Other survey characteristics	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	3	4	0	0	0	0	0
Professionalism	0	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	4	24	0	0	0	0
Recruiter	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	12	3	0	0
Recruitment	0	0	0	7	0	0	0	0	0	0	0	0	0	0	1	2	0	0	7	68	0	0
Routine	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	74	0
Tell opinion	0	0	0	0	0	0	2	0	1	1	0	0	0	0	0	33	0	0	0	0	0	228
NA	0	0	0	0	0	1	0	1	2	0	0	1	0	0	231	3	1	0	0	0	0	0

Table A3: Confusion matrix (actual vs. predicted categories) for GPT under zero-shot prompting with descriptions.

GPT – few-shot

R: Prediction / C: Reference	Anonymity	Brevity	Curiosity	Dutifulness	Fun	Help politicians	Help science	Help society	Help, not further specified	Importance in general	Incentive	Influence	Interest	Learning	No reason	Non-identifiable/Other	Other survey characteristics	Professionalism	Recruiter	Recruitment	Routine	Tell opinion
Anonymity	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Brevity	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0
Curiosity	0	0	38	0	0	0	0	0	0	0	0	0	5	2	2	9	0	0	0	0	0	0
Dutifulness	0	0	0	46	0	0	0	1	1	0	0	0	0	0	0	4	0	0	0	0	2	0
Fun	0	0	0	0	376	0	0	0	0	0	0	0	1	0	0	2	2	0	0	0	0	0
Help politicians	0	0	0	0	0	21	1	2	0	0	0	4	0	0	0	11	0	0	0	0	0	0
Help science	0	0	0	0	0	0	81	1	9	11	0	2	1	0	0	32	0	1	0	0	0	0
Help society	0	0	0	0	0	0	0	99	8	4	0	7	1	0	0	36	0	0	0	0	0	0
Help, not further specified	0	0	0	0	0	0	0	0	24	0	0	1	0	0	0	3	0	0	0	0	0	0
Importance in general	0	0	0	0	0	0	2	3	0	44	0	0	0	0	0	15	5	0	0	0	0	0
Incentive	0	0	0	5	0	0	0	0	0	0	1523	0	0	0	0	3	0	0	1	0	0	0
Influence	0	0	0	0	0	1	0	1	0	1	0	43	1	0	1	15	0	0	0	0	0	2
Interest	0	0	0	0	0	0	0	1	0	0	0	0	1193	3	0	21	27	0	0	0	0	0
Learning	0	0	0	0	0	0	3	2	0	3	0	0	1	28	0	33	0	0	0	0	0	1
No reason	0	0	0	0	0	0	0	0	0	0	250	0	2	0	50	3	4	0	0	0	0	0
Non-identifiable/Other	0	0	0	1	0	0	0	0	0	0	24	0	0	1	20	29	1	0	0	0	0	0

Other survey characteristics	0	0	0	0	0	0	0	0	0	0	0	0	0	46	0	0	7	11	0	0	0	0	0
Professionalism	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	6	23	0	0	0	0
Recruiter	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	11	3	0	0	0
Recruitment	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	68	1	0	0
Routine	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	71	0	0
Tell opinion	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	29	0	0	0	0	0	0	229
NA	0	0	0	0	0	0	0	1	0	0	0	0	0	0	284	2	0	0	0	0	0	0	0

Table A4: Confusion matrix (actual vs. predicted categories) for GPT under few-shot prompting.

GPT – fine-tuned

R: Prediction / C: Reference	Anonymity	Brevity	Curiosity	Dutifulness	Fun	Help politicians	Help science	Help society	Help, not further specified	Importance in general	Incentive	Influence	Interest	Learning	No reason	Non-identifiable/Other	Other survey characteristics	Professionalism	Recruiter	Recruitment	Routine	Tell opinion
Anonymity	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Brevity	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Curiosity	0	0	8	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
Dutifulness	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fun	0	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Help politicians	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0
Help science	0	0	0	0	0	0	11	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
Help society	0	0	0	0	0	0	0	19	0	1	0	1	0	0	0	0	0	0	0	0	0	0
Help, not further specified	0	0	0	0	0	0	0	0	8	0	0	1	0	0	0	0	0	0	0	0	0	0
Importance in general	0	0	0	0	0	0	4	1	0	12	0	0	0	0	0	0	2	0	0	0	0	0
Incentive	0	0	0	1	0	0	0	0	0	0	360	0	0	0	0	0	0	0	0	0	0	0
Influence	0	0	0	0	0	0	0	1	0	0	0	10	0	0	0	3	0	0	0	0	0	0
Interest	0	0	0	0	0	0	1	1	0	0	0	0	248	0	1	1	0	0	1	0	0	0
Learning	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	1	0	0	0	0	0	0
No reason	0	0	0	0	0	0	0	0	0	0	0	0	0	0	63	1	0	0	0	0	0	0
Non-identifiable/Other	0	0	0	0	0	0	2	1	1	0	0	0	1	1	4	34	1	0	0	0	0	0

Other survey characteristics	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	10	1	0	0	0	0
Professionalism	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0
Recruiter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0
Recruitment	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	14	0	0
Routine	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	15	0
Tell opinion	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47
NA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0

Table A5: Confusion matrix (actual vs. predicted categories) for fine-tuned GPT with zero-shot prompt.

Llama – zero-shot

R: Prediction / C: Reference	Anon ymity	Brevit y	Curio sity	Dutifu lness	Fun	Help politic ians	Help scien ce	Help societ y	Help, not furthe r specif ied	Impor tance in gener al	Incent ive	Influe nce	Intere st	Learni ng	No reaso n	Non-i dentifi able/O ther	Other surve y chara cterist ics	Profe ssion alism	Recru iter	Recru itment	Routi ne	Tell opinio n
Anonymity	25	0	0	0	0	0	0	0	0	0	0	0	0	6	15	3	1	0	0	0	0	1
Brevity	0	3	0	1	0	0	0	0	0	0	15	0	0	0	7	2	3	0	0	4	2	0
Curiosity	0	0	34	0	0	2	2	1	0	2	26	2	287	8	23	18	6	0	1	0	0	7
Dutifulness	0	0	0	22	0	1	0	3	1	1	0	6	0	1	0	12	1	1	0	1	0	17
Fun	0	0	0	1	262	0	0	0	0	1	7	0	2	0	0	4	0	0	1	1	0	0
Help politicians	0	0	0	0	0	7	0	4	0	0	0	2	1	0	0	4	0	0	0	0	0	1
Help science	0	0	0	0	0	2	76	27	15	35	3	18	15	10	0	84	7	8	0	0	0	14
Help society	0	0	0	0	0	2	0	38	5	0	4	2	1	0	1	11	1	0	0	0	0	3
Help, not further specified	0	0	0	0	0	0	0	4	10	2	3	5	0	0	0	4	0	1	0	0	0	1
Importance in general	0	0	0	0	0	0	0	5	0	3	0	3	1	1	0	3	2	0	0	0	0	5
Incentive	0	0	0	2	0	0	0	0	0	0	1419	2	0	0	13	6	5	0	1	1	0	6
Influence	0	0	0	0	0	1	0	1	0	1	2	4	0	0	1	2	0	0	0	0	0	6
Interest	0	0	1	0	1	0	0	3	1	5	29	2	643	2	1	15	13	1	2	0	0	23
Learning	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	4	1	0

No reason	0	1	0	6	0	0	0	0	0	0	11	0	2	0	47	5	3	0	0	3	0	2
Non-identifiable/Other	0	0	0	1	1	0	0	0	1	1	13	0	3	1	32	18	2	0	2	0	0	3
Other survey characteristics	0	0	0	0	0	0	0	1	0	1	5	0	0	1	24	5	1	0	0	0	0	11
Professionalism	0	0	0	0	0	1	1	7	1	6	2	3	0	0	0	7	2	7	0	0	1	22
Recruiter	0	0	0	4	0	1	1	0	1	0	28	0	0	0	1	3	2	2	6	14	0	1
Recruitment	0	0	0	8	0	2	5	0	1	1	43	4	0	0	1	9	2	3	6	30	1	2
Routine	0	1	2	7	4	1	0	2	0	0	58	2	0	0	10	13	7	0	1	7	68	1
Tell opinion	0	0	0	0	0	0	1	1	2	1	0	0	0	0	0	3	0	0	0	0	0	80
NA	2	3	1	2	108	2	2	14	4	3	129	2	297	4	183	24	7	1	0	6	1	26

Table A6: Confusion matrix (actual vs. predicted categories) for Llama under zero-shot prompting.

Llama – with descriptions

R: Prediction / C: Reference	Anonymity	Brevity	Curiosity	Dutifulness	Fun	Help politicians	Help science	Help society	Help, not further specified	Importance in general	Incentive	Influence	Interest	Learning	No reason	Non-identifiable/Other	Other survey characteristics	Professionalism	Recruiter	Recruitment	Routine	Tell opinion
Anonymity	27	0	0	0	0	0	1	0	0	0	2	0	0	0	0	3	1	0	0	0	0	0
Brevity	0	7	0	0	0	0	0	0	0	0	0	1	0	0	0	1	9	0	0	0	0	0
Curiosity	0	0	38	0	0	0	5	0	0	2	21	1	39	5	24	20	2	0	0	0	0	4
Dutifulness	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0
Fun	0	0	0	0	375	0	0	0	1	0	2	0	2	4	1	2	3	0	0	0	0	0
Help politicians	0	0	0	0	0	8	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	1
Help science	0	0	0	0	0	1	36	3	4	4	42	3	0	0	61	9	0	0	0	0	0	0
Help society	0	0	0	2	0	3	1	63	3	8	37	7	0	0	32	22	1	0	0	0	0	0
Help, not further specified	0	0	0	1	0	2	2	2	20	2	5	4	0	2	0	9	0	0	0	0	0	0
Importance in general	0	0	0	0	0	0	6	26	2	37	1	0	1	2	0	11	2	0	0	0	0	1
Incentive	0	0	0	6	0	1	0	0	0	0	1478	0	0	0	0	5	0	0	1	0	0	0
Influence	0	0	0	0	0	1	0	0	0	0	2	23	0	0	0	7	0	0	0	0	0	1
Interest	0	0	0	0	0	3	15	8	1	3	24	3	1201	13	10	62	28	0	1	0	0	6
Learning	0	0	0	0	0	0	8	1	0	1	3	1	0	4	3	7	0	0	0	0	0	1
No reason	0	1	0	0	0	0	0	2	1	0	6	2	0	0	19	6	1	0	0	0	0	4

Non-identifiable/Other	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	4	3	0	2	1	0	0
Other survey characteristics	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0
Professionalism	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	1	3	23	0	0	0	0
Recruiter	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	2	0	0	10	1	0	0
Recruitment	0	0	0	10	1	0	0	2	5	0	29	5	0	1	17	17	4	0	6	67	3	0	0
Routine	0	0	0	2	0	0	0	0	0	0	5	0	0	1	7	1	1	0	0	0	66	0	0
Tell opinion	0	0	0	0	0	3	4	3	2	3	63	4	0	2	88	45	3	0	0	0	0	209	0
NA	0	0	0	3	0	0	8	1	1	3	76	2	7	0	95	12	3	1	0	2	5	5	5

Table A7: Confusion matrix (actual vs. predicted categories) for Llama under zero-shot prompting with descriptions.

Llama – few-shot

R: Prediction / C: Reference	Anonymity	Brevity	Curiosity	Dutifulness	Fun	Help politicians	Help science	Help society	Help, not further specified	Importance in general	Incentive	Influence	Interest	Learning	No reason	Non-identifiable/Other	Other survey characteristics	Professionalism	Recruiter	Recruitment	Routine	Tell opinion
Anonymity	27	0	0	0	0	0	0	0	0	0	2	0	0	0	7	0	0	0	1	0	0	0
Brevity	0	7	0	0	0	0	0	0	0	0	9	0	0	0	3	0	8	0	0	2	0	0
Curiosity	0	0	38	0	1	0	0	0	0	0	58	2	16	15	59	11	3	0	0	0	0	2
Dutifulness	0	0	0	35	0	0	0	1	1	0	1	2	0	0	3	4	0	0	0	0	0	0
Fun	0	0	0	0	374	0	0	0	0	0	14	0	5	0	3	2	0	0	0	0	0	1
Help politicians	0	0	0	0	0	14	0	0	0	0	0	3	2	0	1	4	0	0	0	0	0	1
Help science	0	0	0	0	0	2	80	4	8	15	0	4	2	3	6	57	3	2	0	0	0	3
Help society	0	0	0	0	0	3	2	94	6	19	1	13	6	2	8	65	3	0	0	0	0	3
Help, not further specified	0	0	0	0	0	0	0	0	21	1	3	2	0	0	2	4	0	0	0	0	0	0
Importance in general	0	0	0	1	0	0	0	7	1	22	0	3	0	0	6	6	0	0	0	0	0	2
Incentive	0	0	0	4	0	0	1	0	0	0	1505	0	0	0	2	2	0	0	1	0	0	0
Influence	0	0	0	0	0	2	1	0	0	0	1	21	2	1	5	9	3	0	0	0	0	4
Interest	0	0	0	0	0	0	0	1	1	1	36	0	1107	1	3	10	19	0	0	1	0	10
Learning	0	0	0	0	0	0	2	0	0	0	0	0	0	4	3	4	0	1	0	0	0	1
No reason	0	1	0	2	0	0	0	0	0	1	36	0	0	0	43	6	3	0	0	0	0	0
Non-identifiable/Other	0	0	0	1	0	0	0	0	0	0	25	1	3	4	20	15	4	0	1	0	0	1
Other survey	0	0	0	0	0	0	0	0	1	0	3	1	100	1	17	15	7	0	0	0	0	0

characteristics																							
Professionalism	0	0	0	1	0	0	0	1	1	4	2	1	0	2	3	11	5	19	1	0	0	2	
Recruiter	0	0	0	3	0	0	1	0	0	0	2	0	0	0	5	3	0	2	12	30	0	0	
Recruitment	0	0	0	3	0	0	0	0	0	0	1	0	0	0	4	1	0	0	2	34	0	1	
Routine	0	0	0	6	0	0	0	2	0	0	93	0	0	0	17	7	6	0	1	1	74	0	
Tell opinion	0	0	0	0	0	0	0	1	2	0	0	4	0	0	5	15	1	0	0	0	0	201	
NA	0	0	0	0	1	1	1	1	0	0	5	0	9	1	134	4	0	0	1	3	0	0	

Table A8: Confusion matrix (actual vs. predicted categories) for Llama under few-shot prompting.

Mistral – zero-shot

R: Prediction / C: Reference	Anonymity	Brevity	Curiosity	Dutifulness	Fun	Help politicians	Help science	Help society	Help, not further specified	Importance in general	Incentive	Influence	Interest	Learning	No reason	Non-identifiable/Other	Other survey characteristics	Professionalism	Recruiter	Recruitment	Routine	Tell opinion
Anonymity	27	0	0	2	0	0	0	4	1	2	21	1	1	1	1	8	3	0	1	0	0	0
Brevity	0	4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	7	0	0	1	0	0
Curiosity	0	0	31	1	1	3	4	9	2	7	40	3	68	6	12	35	10	0	0	1	0	1
Dutifulness	0	0	0	31	0	0	0	0	0	0	0	1	0	0	1	2	1	1	0	0	0	0
Fun	0	0	0	0	372	0	0	0	0	0	7	0	3	1	0	4	5	0	1	0	0	0
Help politicians	0	0	0	0	0	5	0	0	0	0	0	0	1	0	10	3	0	0	0	0	0	0
Help science	0	0	0	0	0	3	52	4	5	6	2	0	3	1	1	26	0	2	0	0	0	1
Help society	0	0	0	2	0	1	0	46	4	6	3	4	0	0	1	23	0	0	0	0	0	0
Help, not further specified	0	0	0	1	0	2	0	1	9	0	1	6	0	1	0	2	1	0	1	1	0	0
Importance in general	0	0	0	0	0	1	1	9	3	8	5	4	0	4	0	6	3	0	0	0	0	1
Incentive	0	0	0	6	0	0	2	0	0	0	1087	1	0	0	0	5	0	0	1	0	0	0
Influence	0	0	0	1	0	2	0	1	0	1	5	19	0	0	3	11	0	0	0	0	0	2
Interest	0	0	2	1	1	2	7	5	3	11	74	1	1049	7	9	41	18	0	3	0	0	2
Learning	0	0	0	1	0	0	7	4	1	0	11	1	2	8	0	10	1	0	0	1	0	1
No reason	0	1	0	0	0	0	0	1	0	0	2	0	0	0	12	3	1	0	1	2	0	0
Non-identifiable/Other	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	7	0	0	1	0	0	0
Other survey characteristics	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	0	0

Professionalism	0	0	0	0	0	0	2	3	0	3	8	4	1	0	0	6	1	17	1	0	0	0
Recruiter	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	6	27	0	0
Recruitment	0	0	0	1	0	1	0	0	1	0	14	1	0	0	0	7	3	0	2	4	3	0
Routine	0	0	0	1	0	0	0	0	0	0	4	0	0	0	0	1	1	0	0	2	61	0
Tell opinion	0	0	0	0	0	1	2	6	6	8	20	5	2	1	298	28	2	2	0	0	0	199
NA	0	3	5	7	2	1	11	19	7	11	493	5	121	4	8	27	6	2	2	31	10	25

Table A9: Confusion matrix (actual vs. predicted categories) for Mistral under zero-shot prompting.

Mistral – with descriptions

R: Prediction / C: Reference	Anonymity	Brevity	Curiosity	Dutifulness	Fun	Help politicians	Help science	Help society	Help, not further specified	Importance in general	Incentive	Influence	Interest	Learning	No reason	Non-identifiable/Other	Other survey characteristics	Professionalism	Recruiter	Recruitment	Routine	Tell opinion
Anonymity	27	0	0	0	0	0	1	0	0	0	2	0	0	0	0	3	1	0	0	0	0	0
Brevity	0	7	0	0	0	0	0	0	0	0	0	1	0	0	0	1	9	0	0	0	0	0
Curiosity	0	0	38	0	0	0	5	0	0	2	21	1	39	5	24	20	2	0	0	0	0	4
Dutifulness	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0
Fun	0	0	0	0	375	0	0	0	1	0	2	0	2	4	1	2	3	0	0	0	0	0
Help politicians	0	0	0	0	0	8	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	1
Help science	0	0	0	0	0	1	36	3	4	4	42	3	0	0	61	9	0	0	0	0	0	0
Help society	0	0	0	2	0	3	1	63	3	8	37	7	0	0	32	22	1	0	0	0	0	0
Help, not further specified	0	0	0	1	0	2	2	2	20	2	5	4	0	2	0	9	0	0	0	0	0	0
Importance in general	0	0	0	0	0	0	6	26	2	37	1	0	1	2	0	11	2	0	0	0	0	1
Incentive	0	0	0	6	0	1	0	0	0	0	1478	0	0	0	0	5	0	0	1	0	0	0
Influence	0	0	0	0	0	1	0	0	0	0	2	23	0	0	0	7	0	0	0	0	0	1
Interest	0	0	0	0	0	3	15	8	1	3	24	3	1201	13	10	62	28	0	1	0	0	6
Learning	0	0	0	0	0	0	8	1	0	1	3	1	0	4	3	7	0	0	0	0	0	1
No reason	0	1	0	0	0	0	0	2	1	0	6	2	0	0	19	6	1	0	0	0	0	4
Non-identifiable/Other	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	4	3	0	2	1	0	0
Other survey	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0

characteristics																							
Professionalism	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	1	3	23	0	0	0	0	
Recruiter	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	2	0	0	10	1	0	0	
Recruitment	0	0	0	10	1	0	0	2	5	0	29	5	0	1	17	17	4	0	6	67	3	0	
Routine	0	0	0	2	0	0	0	0	0	0	5	0	0	1	7	1	1	0	0	0	66	0	
Tell opinion	0	0	0	0	0	3	4	3	2	3	63	4	0	2	88	45	3	0	0	0	0	209	
NA	0	0	0	3	0	0	8	1	1	3	76	2	7	0	95	12	3	1	0	2	5	5	

Table A10: Confusion matrix (actual vs. predicted categories) for Mistral under zero-shot prompting with descriptions.

Mistral – few-shot

R: Prediction / C: Reference	Anonymity	Brevity	Curiosity	Dutifulness	Fun	Help politicians	Help science	Help society	Help, not further specified	Importance in general	Incentive	Influence	Interest	Learning	No reason	Non-identifiable/Other	Other survey characteristics	Professionalism	Recruiter	Recruitment	Routine	Tell opinion
Anonymity	27	0	0	0	0	0	0	0	0	0	7	0	0	0	1	3	0	0	0	0	0	0
Brevity	0	8	0	0	0	0	0	0	0	0	0	0	0	0	6	1	11	0	0	0	0	0
Curiosity	0	0	38	0	1	0	3	0	0	1	17	0	15	10	4	9	0	0	0	0	0	0
Dutifulness	0	0	0	33	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0
Fun	0	0	0	0	375	0	0	0	0	0	0	0	2	0	0	0	5	0	0	0	0	0
Help politicians	0	0	0	0	0	12	0	0	0	0	1	1	1	0	0	2	0	0	0	0	0	0
Help science	0	0	0	0	0	3	74	3	7	12	2	1	2	0	1	49	1	1	0	0	0	0
Help society	0	0	0	0	0	4	0	94	10	6	2	12	2	1	35	44	1	0	0	0	0	0
Help, not further specified	0	0	0	2	0	0	4	6	21	0	18	1	0	0	12	5	0	0	0	0	0	0
Importance in general	0	0	0	0	0	0	1	6	0	42	9	2	0	1	0	12	3	0	0	0	0	0
Incentive	0	0	0	7	0	0	0	0	0	0	1572	0	0	0	18	7	0	0	1	0	0	0
Influence	0	0	0	1	0	3	0	0	0	0	1	31	0	0	0	10	0	0	0	0	0	2
Interest	0	0	0	0	0	0	0	1	0	0	28	0	1162	5	5	28	25	0	0	0	0	0
Learning	0	0	0	0	0	0	2	1	0	0	0	0	0	8	4	8	0	0	0	0	0	0
No reason	0	0	0	0	0	0	0	0	0	0	2	1	0	0	19	9	3	0	0	0	0	0
Non-identifiable/Other	0	0	0	0	0	0	0	0	0	0	28	0	2	3	50	4	1	0	0	0	0	0
Other survey characteristics	0	0	0	0	0	0	0	0	0	0	2	0	62	2	20	13	9	0	0	0	0	0

Professionalism	0	0	0	0	0	0	2	0	0	0	2	0	0	0	0	1	4	23	0	0	0	0
Recruiter	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	7	2	0	0
Recruitment	0	0	0	13	0	0	0	0	3	0	95	4	1	0	0	11	2	0	10	69	2	1
Routine	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	3	0	0	0	0	72	0
Tell opinion	0	0	0	0	0	0	2	0	1	2	3	3	1	4	147	32	0	0	0	0	0	227
NA	0	0	0	0	0	0	0	0	0	0	5	1	2	0	37	2	0	0	2	0	0	2

Table A11: Confusion matrix (actual vs. predicted categories) for Mistral under few-shot prompting.

6. Fine-Tuning Performance

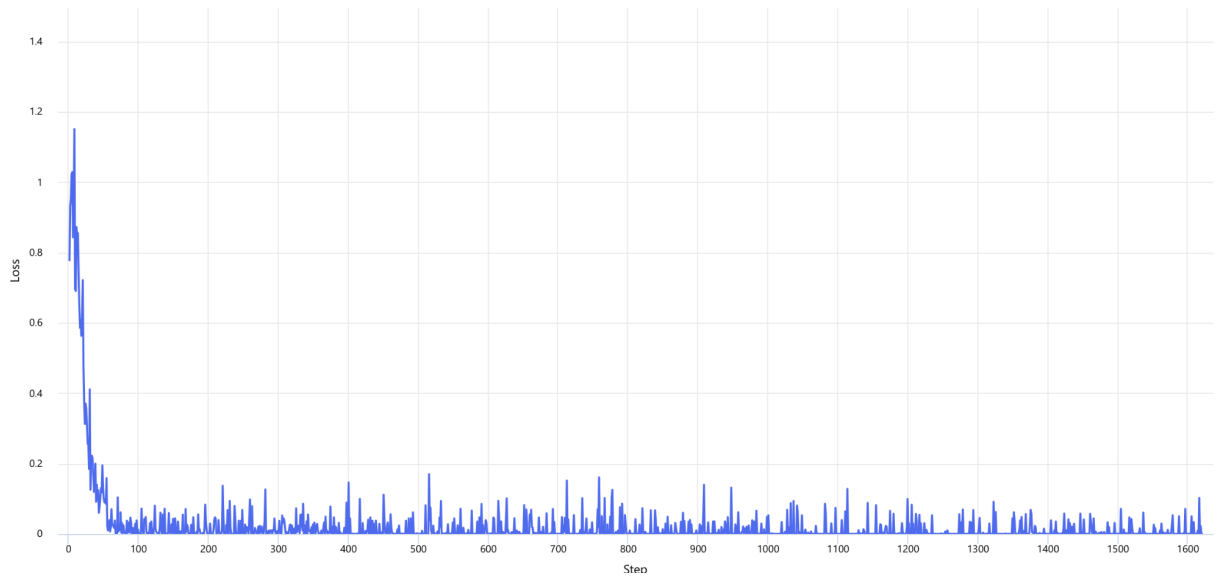


Figure A7: Loss curve during fine-tuning.

*Note: $\text{steps} = n_{\text{examples}} / \text{batch_size} * n_{\text{epoch}}$. Epochs describe the number of iterations through the data, and batch size the number of examples used in a single training pass.*

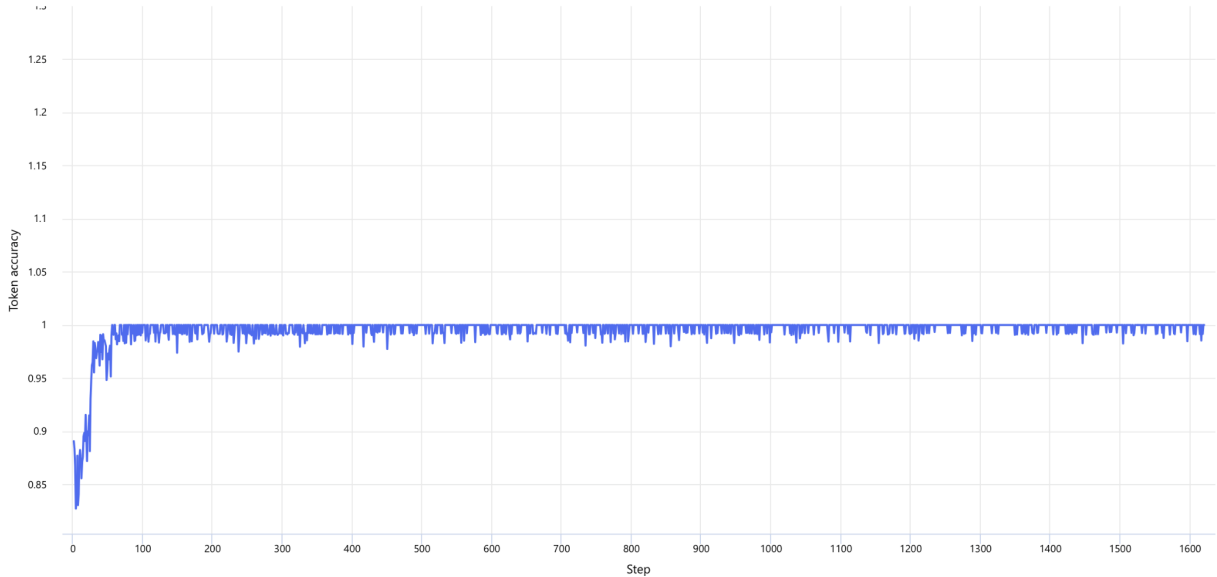


Figure A8: Mean token accuracy achieved during fine-tuning.