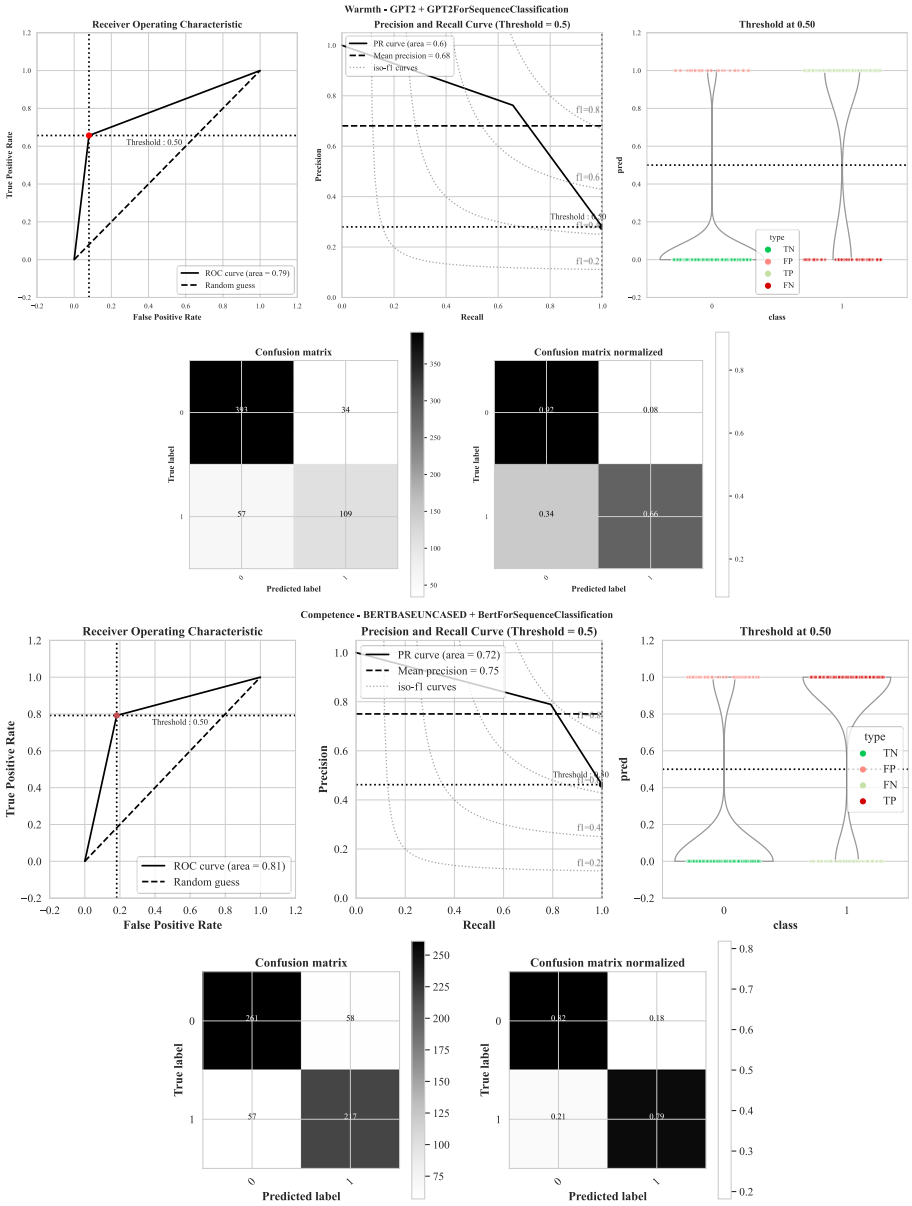
1 Supplementary Material

2 Appendix A

Figure A1

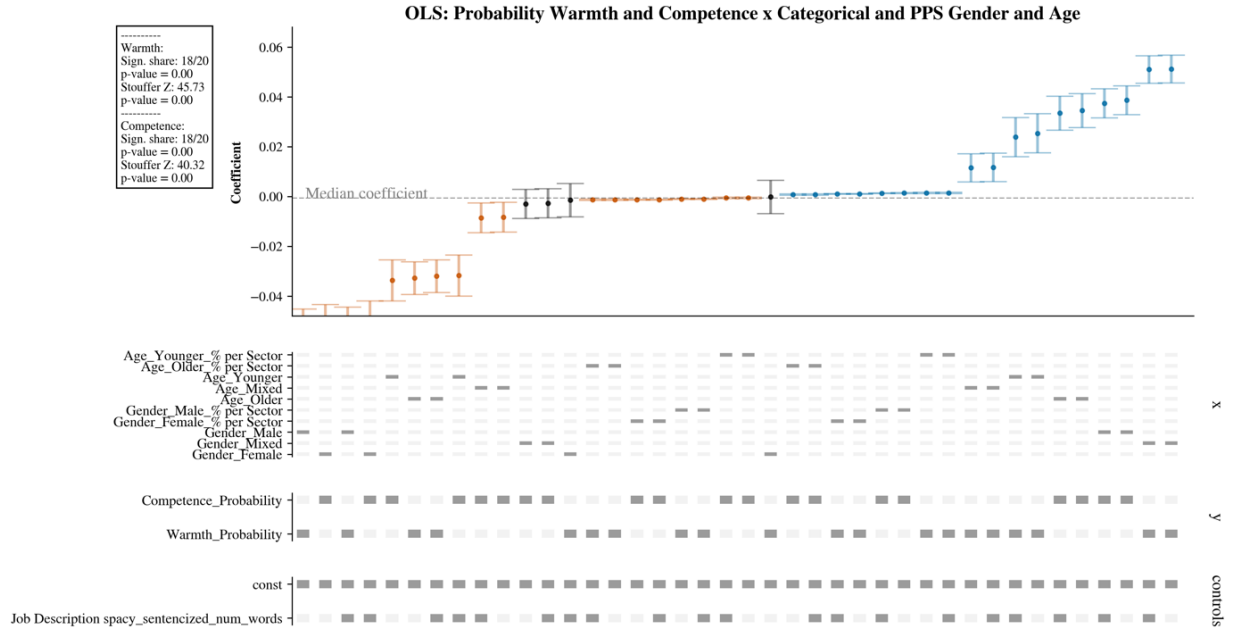
Classifying warmth- and competence-related framing: ROC curve, precision-recall curve, and confusion matrices of true and predicted results obtained from GPT2 finetuned for warmth and BERT model finetuned for competence



Note. 1 indicates datapoints labeled as related to warmth or competence respectively and 0 indicates datapoints labeled as not related to warmth or competence respectively.

Figure A2

Overall OLS Regression-based Specification Curve on Aggregate Data



Note. Dependent variable: Probability of Warmth- and Competence-Related Frames Presence in Sentence (PPW and PPC respectively). Independent variables: (1) Categorical Dominant Social Group of Sector and (2) Percentages of Social Group per Sector (PPS). Control variable: Number of Words per Sentence

Table A1

Warmth multiple regression model (Full Model) and specification curve analysis (SCA) regression results

	Warmth Full Model	Warmth Model 1	Warmth Model 2	Warmth Model 3	Warmth Model 4	Warmth Model 5	Warmth Model 6	Warmth Model 7	Warmth Model 8	Warmth Model 9	Warmth Model 10
Gender_Female	-0.02***								0.01***		
	(0.00)								(0.00)		
Gender_Mixed	-										0.03***
	-										(0.00)
Gender_Male	-0.02***	-0.04***									
	(0.00)	(0.00)									
Gender_Female_% per Sector	-0.12***					0.00***					
	(0.01)					(0.00)					
Gender_Male_% per Sector	-0.12***			-0.00***							
	(0.01)			(0.00)							
Age_Older	-0.04***		-0.03***								
	(0.00)		(0.00)								
Age_Mixed	-							0.01***			
	-							(0.00)			
Age_Younger	0.00									0.02***	
	(0.00)									(0.00)	
Age_Older_% per Sector	0.02***				-0.00***						
	(0.00)				(0.00)						
Age_Younger_% per Sector	0.02***						0.00***				
	(0.00)						(0.00)				
Job Description num_words	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
const	10.19***	0.20***	0.19***	0.24***	0.22***	0.15***	0.13***	0.18***	0.18***	0.18***	0.18***
	(1.12)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
F	5158.33	22781.13	22417.70	22630.06	22294.16	22627.40	22306.53	22227.26	22232.60	22239.54	22490.60
F (p-value)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

df_model	9	2	2	2	2	2	2	2	2	2	2
df_resid	308573	308580	308580	308580	308580	308580	308580	308580	308580	308580	308580
R-squared	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
R-squared Adj.	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
Unstandardized Coefficent B (b)		-0.04	-0.03	-0.00	-0.00	0.00	0.00	0.01	0.01	0.02	0.03
Standard Error (SE)		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Standardized Coefficient b* (β)		-0.11	-0.09	-0.00	-0.00	0.00	0.00	0.03	0.04	0.05	0.08
t		-32.49	-20.50	-28.14	-14.30	28.06	15.03	9.35	9.84	10.43	23.41
t (p-value)		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
95% CI		-0.04 - -0.04	-0.03 - -0.03	-0.00 - -0.00	-0.00 - -0.00	0.00 - 0.00	0.00 - 0.00	0.01 - 0.01	0.02 - 0.02	0.02 - 0.02	0.03 - 0.03

Note.
Models are ordered by independent variable type.
Standard errors in parentheses.
** p<.1, ** p<.05, ***p<.01*

Table A2

Competence multiple regression model (Full Model) and specification curve analysis (SCA) regression models and results

	Comp Full Model	Comp Model 1	Comp Model 2	Comp Model 3	Comp Model 4	Comp Model 5	Comp Model 6	Comp Model 7	Comp Model 8	Comp Model 9	Comp Model 10
Gender_Female	-0.06***		-0.02***								
	(0.00)		(0.00)								
Gender_Mixed	-				-0.01***						
	-				(0.00)						
Gender_Male	-0.00									0.03***	
	(0.00)									(0.00)	
Gender_Female_% per Sector	0.32***					-0.00***					
	(0.01)					(0.00)					
Gender_Male_% per Sector	0.32***								0.00***		
	(0.01)								(0.00)		
Age_Older	0.01***										0.03***

	(0.00)										(0.00)
Age_Mixed	-			-0.01***							
	-			(0.00)							
Age_Younger	-0.03***	-0.03***									
	(0.00)	(0.00)									
Age_Older_% per Sector	0.01***							0.00***			
	(0.00)							(0.00)			
Age_Younger_% per Sector	0.01***						-0.00***				
	(0.00)						(0.00)				
Job Description num_words	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
const	- 32.44***	0.33***	0.33***	0.33***	0.32***	0.36***	0.37***	0.29***	0.28***	0.31***	0.32***
	(1.17)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
F	4041.81	17158.57	17191.99	17058.73	17052.80	17423.32	17115.96	17138.26	17426.87	17258.08	17263.84
F (p-value)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
df_model	9	2	2	2	2	2	2	2	2	2	2
df_resid	308573	308580	308580	308580	308580	308580	308580	308580	308580	308580	308580
R-squared	0.11	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
R-squared Adj.	0.11	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Unstandardized Coefficent B (b)		-0.03	-0.02	-0.01	-0.01	-0.00	-0.00	0.00	0.00	0.03	0.03
Standard Error (SE)		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Standardized Coefficient b* (β)		-0.07	-0.07	-0.02	-0.02	-0.00	-0.00	0.00	0.00	0.07	0.09
t		-14.63	-16.56	-5.84	-4.84	-26.28	-11.71	13.32	26.40	19.83	20.09
t (p-value)		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
95% CI		-0.02 - - 0.02	-0.02 - - 0.02	-0.01 - - 0.01	-0.00 - - 0.00	-0.00 - - 0.00	-0.00 - - 0.00	0.00 - 0.00	0.00 - 0.00	0.03 - 0.03	0.04 - 0.04

Note.
Models are ordered by independent variable type.
Standard errors in parentheses.
** p<.1, ** p<.05, ***p<.01*

Table A3

Sectoral Gender and Age Composition and Percentages

Jobs Count per Sector (x 1000)											
SBI Sector Titles		Gender					Age				
Industry class / branch (SIC2008)		Female		Male		Sectoral Gender Segregation	Older (>=45 years)		Younger (< 45 years)		Sectoral Age Segregation
Code	Sector Name	*n*	% per Sector	*n*	% per Sector	Dominant Category	*n*	% per Sector	*n*	% per Sector	Dominant Category
A-U	All economic activities	4029		4362			3500		4892		
A	Agriculture and industry	310	22.2	1089	77.8	Male	690	49.3	708	50.6	Mixed Age
B	Industry (no construction), energy	423	20.9	1599	79.1	Male	1066	52.7	954	47.2	Older
C	Manufacturing	174	22.7	592	77.3	Male	413	53.9	354	46.2	Older
D	Energy supply	8	27.6	21	72.4	Male	15	51.7	13	44.8	Mixed Age
E	Water supply and waste management	7	19.4	29	80.6	Male	21	58.3	16	44.4	Older
F	Construction	42	12.5	294	87.5	Male	160	47.6	178	53.0	Mixed Age
G	Commercial services	3421	43.1	4510	56.9	Mixed Gender	2704	34.1	5228	65.9	Mixed Age
H	Transportation and storage	95	24.6	291	75.4	Male	205	53.1	181	46.9	Older
I	Accommodation and food serving	199	50.3	196	49.5	Mixed Gender	75	18.9	320	80.8	Younger
J	Information and communication	80	27.6	210	72.4	Male	95	32.8	195	67.2	Mixed Age
K	Financial institutions	108	39.4	166	60.6	Mixed Gender	146	53.3	128	46.7	Older
L	Renting, buying, selling real estate	33	48.5	35	51.5	Mixed Gender	36	52.9	33	48.5	Older
M	Business services	857	42.1	1177	57.9	Mixed Gender	726	35.7	1307	64.3	Mixed Age
N	Renting and other business support	416	42.8	557	57.2	Mixed Gender	311	32.0	660	67.8	Younger
O	Government and care	3970	68.2	1855	31.8	Female	2844	48.8	2981	51.2	Mixed Age
P	Education	353	65.1	189	34.9	Female	253	46.7	288	53.1	Mixed Age

Q	Health and social work activities	1208	84.3	224	15.6	Female	661	46.1	770	53.7	Mixed Age
R	Culture, recreation, other services	226	56.8	171	43.0	Mixed Gender	155	38.9	241	60.6	Mixed Age
S	Other service activities	87	65.9	45	34.1	Female	55	41.7	75	56.8	Mixed Age
	Total (excluding A-U)	12017	47.6	13250	52.4		10631	42.1	14630	57.9	

Note.
Threshold for gender=47.55% ± 20%
Threshold for age=42.07% ± 10%
Source: Centraal Bureau voor de Statistiek (CBS)

3 Appendix B

4 Technical Methodology

4.1 Automated Content Analysis

To automate content analysis, we compared (1) traditional supervised classifiers and (2) pre-trained large language transformer classifiers finetuned for our classification task. We implemented binary classification where each classifier was trained on one dependent variable, i.e., presence (vs. absence) of warmth- and competence-related frames respectively. The training data consisted of 5947 human-annotated sentences split into training, evaluation, and testing datasets (75:15:10). The distribution of positive and negative classes was imbalanced for warmth; warmth $n_{sentences}$ present=1615 (27.2%), $n_{sentences}$ absent=4332 (72.8%), M=0.27, SD=0.44, imbalance ratio (IR)=0.37. The imbalance ratio (IR) for competence was 0.87. Note that the closer to 1 the IR, the more balanced a dataset is considered (Zhu et al., 2020).[1] We addressed imbalance as automated classifiers "learn" by exposure to negative (absent) and positive (present) instances of a class. If instances of a positive class relative to negative, or vice versa, are limited due to imbalance, classifiers tend to underperform in classifying said class instance, notwithstanding the imbalanced distribution occurring naturally. Given that our sample size is smaller than the usual sample for automated classification, imbalance may have a larger impact. We thus considered the warmth imbalance (but not competence) to be relatively strong and took measures to correct it.

4.2 Model Training

For traditional supervised classification, the estimators were fed preprocessed text data (without stop-words, capitalization, and punctuations, tokenized, and allowed a 1-3 n-gram range). Feature representation was done via a count vectorizer and a term frequency-

inverse document frequency (TF-IDF) vectorizer, and, where feasible, a concatenated vectorizer combining the two. As the distribution of positive and negative classes was imbalanced for warmth, we implemented repeated-stratified 10-fold cross-validation and minority class resampling via SMOTE (Aurelio et al., 2022; Sayyed, 2021).

For the transformers, unprocessed data were fed to BERT base uncased, GPT2, and OpenAI GPT models for sequence classification and their respective fast tokenizers. The models were finetuned via Huggingface's Trainer class; train epochs=3, weight decay=0.01, learning rate=5e-5 (optimized via AdamW). Training evaluation was set to 500 steps and evaluation was done per step. To remedy discrepant predictions particular to binary classification and class imbalance, a custom loss function was implemented. The function used binary cross entropy with a sigmoid activation layer that penalized binary misclassification more rigidly. The function also used positive class weight to account for imbalance.

4.3 Final Model

The models were evaluated on the held-out testing dataset and selection criteria were based on recall of the positive class, Matthews correlation coefficient (MCC), area-under-the-curve (AUC), and precision-recall curve (Burscher et al., 2014; Chicco & Jurman, 2020; Seliya et al., 2009). Additional considerations included whether a given combination performed better than a baseline dummy classifier and a Naïve Bayes model. An overview of the models used and their performance metrics for warmth and competence respectively are provided in the online repository.

For warmth-related frame classification, the GPT2 model had better overall results on validation; recall=0.66, MCC=0.61, precision=0.76, accuracy=0.79, f1-score=0.71, AUC=0.91. For competence-related frame classification, the BERT model was selected; recall=0.79, MCC=0.61, precision=0.79, accuracy=0.81, f1-score=0.79, AUC=0.90. To

confirm human-labeled and classifier-predicted values did not differ significantly, an independent samples t-test was conducted between labels and predictions in the test dataset, i.e., the dataset not encountered by classifiers during training. No statistically significant difference was found for warmth values; $t(1181.24)=1.52$, $p=.128$, nor for competence; $t(1183.99)=0.06$, $p=.954$. Furthermore, we confirm a strong linear relationship between the manually coded and classifier-labeled dependent variables via an OLS regression on the evaluation dataset with manually coded labels as dependent variable and binary predictions as independent variable. The test showed a significant strong linear relationship; warmth $F(1, 592)=1908$, $p=.00$, $R^2=0.76$, $b=0.90$, $t=43.68$, $p=.00$, 95%CI[0.859,0.940], and competence $F(1, 592)=10620$, $p=.00$, $R^2=0.94$, $b=0.83$, $t=103.06$, $p=.00$, 95%CI[0.814, 0.846].

In the interest of explainability, we follow advice from Parasurama et al. (2022) and calculate SHAP (Shapley Additive Explanations) values for job ad sentence in the testing dataset; $n=593$.[2] SHAP values indicate the most predictive features (i.e., tokens or words) used by classification models. Due to computational limitations, we compute SHAP values only for the BERT model used to classify competence-related frames where we find the terms "precision", "execute", and "proficiency" are most predictive of competence framing. The online repository contains a table displaying the full list of predictive features for competence-related frames and associated SHAP values for the BERT model.

The trained model was then used to label job ad sentences as not containing warmth/competence-related frames (0) or containing related to warmth/competence-related frames (1) respectively; warmth $n_{sentences}$ present=74217 (24.1%), $n_{sentences}$ absent=234366 (75.9%), $M=0.24$, $SD=0.43$, and competence $n_{sentences}$ present=146596 (47.5%), $n_{sentences}$ absent=161987 (52.5%), $M=0.48$, $SD=0.50$. Furthermore, the model provided sigmoid probabilities for the presence of warmth- and competence-related frames in a job ad sentence,

which were then used for analysis (see Dependent Variable). Figure A1 shows the confusion matrices, precision-recall curve, and ROC curve.

References

Aurelio, Y. S., de Almeida, G. M., de Castro, C. L., & Braga, A. P. (2022). Cost-Sensitive Learning based on Performance Metric for Imbalanced Data. *Neural Processing Letters*. https://doi.org/10.1007/s11063-022-10756-2

Burscher, B., Odijk, D., Vliegenthart, R., de Rijke, M., & de Vreese, C. H. (2014). Teaching the Computer to Code Frames in News: Comparing Two Supervised Machine Learning Approaches to Frame Analysis. *Communication Methods & Measures*, *8*(3), 190–206. https://doi.org/10.1080/19312458.2014.937527

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1), 6. https://doi.org/10.1186/s12864-019-6413-7

Data Preparation and Feature Engineering in Machine Learning: Imbalanced Data. (2022, July 18). [Course]. Google Developers. https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data

Parasurama, P., Sedoc, J., & Ghose, A. (2022). *Gendered Information in Resumes and Hiring Bias: A Predictive Modeling Approach* (SSRN Scholarly Paper 4074976). https://doi.org/10.2139/ssrn.4074976

Sayyed, Z. A. (2021). Study of sampling methods in sentiment analysis of imbalanced data. *arXiv:2106.06673 [Cs]*. http://arxiv.org/abs/2106.06673

Seliya, N., Khoshgoftaar, T. M., & Van Hulse, J. (2009). A Study on the Relationships of Classifier Performance Metrics. *2009 21st IEEE International Conference on Tools with Artificial Intelligence*, 59–66. https://doi.org/10.1109/ICTAI.2009.25

Zhu, R., Guo, Y., & Xue, J.-H. (2020). Adjusting the imbalance ratio by the dimensionality of imbalanced data. *Pattern Recognition Letters*, *133*, 217–223. https://doi.org/10.1016/j.patrec.2020.03.004

[1] There is no set rule of thumb or threshold for classifying data as prohibitively imbalanced, however, an imbalance ratio of 20/40 would be classified as mild imbalance (*Data Preparation and Feature Engineering in Machine Learning: Imbalanced Data*, 2022).
[2] SHAP calculations could not be conducted on a larger sample due to computational limitations.