Discussion Paper

To share or not to share? Text analysis of survey responses to detect motivations for (not) donating smartphone sensor data to Statistics Netherlands

Marc Smeets
Jeldrik Bakker
Vivian Meertens

November 27, 2025

Abstract

Open-ended answers in surveys capture rich motivations but are costly to code by hand. We study respondents' stated reasons for (probable) refusal to share smartphone-sensor data, using two closely related Dutch questionnaires fielded in 2017–2018: a LISS panel study and a Statistics Netherlands consent survey. Both questionnaires included a common core of sensor tasks (GPS location, selfie, photo of the house exterior, short video), while the LISS panel additionally asked about connecting a wearable (fitness bracelet) and the Statistics Netherlands survey included a receipt photo task. The responses in the LISS panel were coded manually, while there are no manual codes available for the consent survey. However, the consent survey data contain background characteristics (e.g., Gender, Age class and Education level) and the answers for the consent survey are also available per sensor task. We transfer an 11-category motivation taxonomy from the LISS panel to the Statistics Netherlands survey via a transparent NLP pipeline: light normalization, rule-based keyword extraction, and elastic-net logistic regression with a single probability threshold for multi-label assignment. The manually coded responses in the LISS panel can serve as training set and offer the possibility to evaluate the responses classified by the NLP pipeline. The cross-validated AUCs applied to the LISS data are high for core categories like Privacy, Safety and Due to emotions. With threshold $d = 0.2$, only 0.5% of Statistics Netherlands answers remain unassigned and 71.8% receive a single label. The results from both surveys show the following reasons for refusal: *privacy* dominates; *control* and brief refusals (*without reason*) follow. On the reasons per task, it follows from the consent survey, that *location* elicits comparatively more *control*/*surveillance*, while the *house photo* is most often flagged as *effortful*. Demographic contrasts are modest but suggestive (e.g., more *effort* among ages 50–67; more *control*/*surveillance* mentions among younger groups). The approach used for coding the short survey texts is scalable while remaining interpretable. For this application, it is assumed that for practical use, the LISS-trained classification model is transferable to the Statistics Netherlands consent survey.

Keywords: text-mining techniques, survey open-ended questions, natural language processing (NLP), smartphone sensor data.

Contents

1 Introduction

1.1 Background

Open-ended questions are very valuable in survey research, yet their analysis typically depends on manual coding, a process that is labor-intensive and sensitive to coder subjectivity. In parallel, the growing use of smartphones and sensors in data collection raises important questions about respondents' willingness to share sensor-based information in surveys, and about why some respondents refuse. Understanding these motivations is essential for survey design, communication, and data-donation strategies at national statistical institutes such as Statistics Netherlands.

This paper focuses on respondents' stated reasons for (probable) refusal to share smartphone-sensor data. To get insight into these reasons, we analyze the responses from two closely related questionnaires fielded in close succession: a study in the LISS (Longitudinal Internet Studies for the Social Sciences) panel in 2017–2018, and a consent survey conducted by Statistics Netherlands in mid-2018. Both questionnaires included a common core of sensor tasks (GPS location, selfie, photo of the house exterior, short video), while the LISS panel additionally asked about connecting a wearable (fitness bracelet) and the Statistics Netherlands survey included a receipt photo task. In each instrument, respondents who indicated that they would probably or definitely not share were prompted to provide an open-ended motivation for that decision.

In the present study, the texted data of the consent survey are analyzed by automatically classifying the open answers, where we assume the same categories of motivations as for the LISS panel. The responses of the consent survey are classified by means of traditional text mining techniques. The most relevant keywords are extracted from the text fields by using natural language processing (NLP) and a rule-based language model specifically designed for this purpose. Subsequently, the probabilities of the motivations are predicted using logistic regression, where the extracted keywords are used as explanatory variables. The manually coded responses in the LISS panel are used to train both models and to evaluate the performance of the classification models. Although no (external) evaluation is possible on the response from the consent survey, due to the lack of manual labels, we still think that the models are of practical use to analyze and interpret the results of the consent survey. Both surveys are closely related, the questionnaires are near-identical, and the studies were conducted in quick succession.

1.2 Related work

The most directly relevant prior work regarding the willingness to share sensor and passive data is by Struminskaya et al. (2020) and Struminskaya et al. (2021), who investigated willingness to share smartphone-sensor measurements in the LISS panel and in the Statistics Netherlands consent survey. Our study builds on the same data sources, extending their analyses by automatically classifying open answers with text-mining techniques. Critically for the present paper, the per-scenario open answers were manually coded into eleven motivation categories (e.g., privacy, perceived misuse, safety, effort, control, surveillance), which we adopt as our taxonomy. Related studies examine participation in and consent to passive or app-based data collection in surveys, highlighting the roles of privacy concerns, trust, perceived benefits, and burden/effort as key determinants (Keusch et al. 2019). Framing and sponsor cues can shift

willingness at the margins, yet concerns about privacy and control remain central across contexts (Struminskaya et al. 2020; Struminskaya et al. 2021).

There is a substantial literature on applying computational text analysis to survey text. For short and domain-specific answers typical of questionnaires, researchers often favor transparent pipelines: light normalization, interpretable features (e.g., dictionaries or rule-based keywording), and supervised models with regularization (Chai 2019; Seni and Elder 2010). Foundational work on text-as-data underscores the need to align methods with substantive goals and to validate against human labels (Grimmer and Stewart 2013; Roberts et al. 2014). For Dutch short texts, part-of-speech tagging and dependency parsing via `udpipe` are widely used to support feature extraction (Straka et al. 2016; Wijffels 2023). Penalized logistic regression (elastic net) provides a pragmatic balance between performance and interpretability in high-dimensional settings (Friedman et al. 2010; Tay et al. 2023).

Even when questionnaires are near-identical, transferring classifiers trained on one study to another can be challenged by shifts in population, item context, or lexical choices. The machine-learning literature treats this under domain adaptation or transfer learning (Daumé III 2007; Pan and Yang 2009). In the survey context, such transfer is particularly attractive for national statistical institutes that field recurring modules and long-running surveys to measure trends over time: a portable, interpretable classifier enables consistent coding across waves, rapid onboarding of new data, and substantial efficiency gains, while safeguarding comparability. At the same time, vocabularies and salience of concerns may evolve; therefore, practical deployments should include light-weight performance monitoring for drift (e.g., rolling validation on small hand-labeled subsamples), periodic recalibration or retraining when needed, and clear governance of the category scheme (versioning and change logs) to maintain continuity of official statistics.

1.3 Research questions

We address the following research questions:

1. **Categories of reasons.** Which motivation categories can be identified in respondents' open answers about not sharing smartphone-sensor data, and how prevalent are these categories overall?
2. **Differences by sensor type.** How do the distributions of motivation categories differ across sensor types (GPS location , selfie, photo of the house, receipt, video)?
3. **Demographic variation.** To what extent are motivation categories associated with demographic characteristics (gender, age group, education level)?

The remainder of the paper is structured as follows. Section 2 describes the two data sources and their harmonization. Section 3 details the NLP pipeline and classification models. Section 4 presents the results organized by the research questions. Section 5 provides a discussion, and conclusion including limitations and implications for practice.

2 Data

This study uses two survey sources that cover similar smartphone-sensor tasks but were fielded in different settings: (i) a study in the LISS panel in 2017–2018, and (ii) a consent survey

conducted by Statistics Netherlands in mid-2018. For each subsection below, we first describe the LISS design and then the Statistics Netherlands consent survey so that the overlap and differences are explicit.

2.1 Instruments and fieldwork

LISS panel study The LISS study investigated one-time willingness to share data from smartphone sensors among panel members in the Netherlands. Respondents were presented with several active sensor tasks (e.g., taking a selfie, a photo of the house exterior, recording a short video), a passive/background task (sharing geolocation/GPS), and were also asked about connecting a fitness bracelet (wearable). The ordering of tasks was randomized where applicable. Whenever a respondent indicated non-willingness for a task, an open-ended prompt invited them to explain their reason(s). Fieldwork took place in late 2017 and spring 2018 in the probability-based LISS panel administered by CentERdata (Struminskaya et al. 2020).

Statistics Netherlands consent survey. This consent survey assessed willingness to donate data collected via smartphones. When a respondent declined a sensor task, they received a follow-up open-ended question to motivate their refusal; a separate final open comment box was also provided at the end of the questionnaire. The sensor tasks were: (1) GPS location, (2) selfie, (3) photo of the outside of the house, (4) photo of a recent purchase receipt, and (5) short video of the surroundings. For camera-based tasks, respondents were transferred to the device's camera app and could review the capture before submission (Groot de 2018; Struminskaya et al. 2021).

2.2 Populations and sampling frames

LISS panel study. The LISS panel comprises approximately 7,000 individuals aged 16+ in private households in the Netherlands, recruited via probability sampling of addresses. For the smartphone-sensor study, invitations targeted 3,023 panel members with a compatible smartphone. At each sampled address, one household member was randomly selected for the advance letter, but all household members were invited to participate once the household joined the panel (Struminskaya et al. 2020).

Statistics Netherlands consent survey. The sampling frame consisted of prior Statistics Netherlands survey respondents of general population samples who completed on a smartphone or tablet and consented to recontact. A stratified sample of 3,200 persons aged 16+ was drawn with strata 16–29, 30–49, and 50+ years. Within each stratum, a systematic sample was selected after sorting by residential address (postal code and house number) (Groot de 2018; Struminskaya et al. 2021).

2.3 Questionnaire design and open-ended questions

LISS panel study. Respondents were asked about their (un)willingness to share data collected with smartphone sensors. For any sensor task where a respondent indicated non-willingness, an open-ended follow-up asked:

– "Could you specify why you would not want to share the data?"

The wording and placement of the open-ended item were consistent across the sensor tasks so that reasons could be compared across tasks (Struminskaya et al. 2020).

Statistics Netherlands consent survey. The questionnaire included attitudinal items about trust in organizations (e.g., Statistics Netherlands, universities, market research companies) and handling of privacy-sensitive data, followed by the sensor tasks. For each task that a respondent declined, the survey asked:

1. "You did not want to or could not [**task description**] for this survey. What are your main reasons for this?"
   a. take a photo of yourself (**selfie**)
   b. share information about your location (**location**)
   c. take a photo of your house (**house**)
   d. take a photo of a receipt (**receipt**)
   e. take a video of your surroundings (**video**)
2. "Finally, do you have any comments about this questionnaire? We would love to hear your opinion." (**final remarks**)

In parentheses we indicate the sensor labels used throughout the paper (Groot de 2018; Struminskaya et al. 2021).

2.4 Background variables

We study heterogeneity in motivations between different groups of respondents based on the response from the *Statistics Netherlands consent survey*. For the *LISS panel study* dataset no background variables linked at the answer level were available. Specifically, we consider gender (male, female), age class (16–29, 30–49, 50–67, 68+), and education level (lower, intermediate, higher). The composition of the Statistics Netherlands consent-survey respondents with respect to these characteristics is given in Appendix A.

2.5 Response and availability of open-text answers

Table 2.1 summarises, by sensor task, (i) the share of respondents who were not willing to share (or who filled in the final remarks) (LISS and consent survey) and, where available, (ii) how often a open-ended reason was provided among those unwilling to share (consent survey only).

For the consent survey, the overall response was 1,965 (61.4%). Willingness differed by task: for GPS location, 50.4% of decliners provided a reason, versus 82–88% for the camera-based tasks; the general 'final remarks' box was filled by 44% of completes.

For the *LISS panel study*, the overall response was 2,673 (88.4%), and 1,818 respondents provided at least one open-ended reason linked to a declined sensor task. Per-task non-willingness ("probably no" + "definitely no") is taken from Struminskaya et al. (2020) and indicates the fraction that could, in principle, supply a motivation if prompted; per-task open-text response fractions are not available for LISS.

Although both surveys were conducted in Dutch, not all respondents answered in Dutch. These answers were translated into Dutch during the data cleaning process (see Subsection 3.1).

Table 2.1 Non-willingness by sensor task and open-text availability.

Survey	Question	Not willing to share	Response	Response fraction
LISS panel	Selfie	2201 (82.3%)	–	–
	Location	1873 (70.1%)	–	–
	House	1652 (61.8%)	–	–
	Video	2042 (76.4%)	–	–
	Wearable	1062 (39.7%)	–	–
Statistics Netherlands consent survey	Selfie	1694 (86.2%)	1399	82.6%
	Location	1100 (56.0%)	554	50.4%
	House	1729 (88.0%)	1505	87.0%
	Receipt	1615 (82.2%)	1355	83.9%
	Video	1656 (84.3%)	1459	88.1%
	Final remarks	–	865	44.0%

2.6 Category prevalence of the manual codes LISS panel

For the *LISS panel study*, the open answers were manually coded into 11 motivation categories by Struminskaya et al. (2020). The categories are listed in Table C.1; multi-label assignments were allowed. Table 2.2 shows how many categories were manually assigned per open answer. Most answers received a single category (81.9%), while 18.1% received multiple labels. The denominator is the set of valid open answers ($n = 1{,}818$); the "Missing" row refers to panel respondents without an open answer.

Table 2.2 LISS panel study: number of manually assigned categories per answer.

	# categories	Frequency	Percent	Valid percent
Valid	1	1489	55.7%	81.9%
	2	303	11.3%	16.7%
	3	24	0.9%	1.3%
	4	2	0.1%	0.1%
	total	1818	68.0%	100.0%
Missing		855	32.0%	
Total		2673	100.0%	

Note: Because multiple categories can be assigned to a single answer, the $n = 1{,}818$ distinct answers correspond to $1 \times 1489 + 2 \times 303 + 3 \times 24 + 4 \times 2 = 2{,}175$ category assignments in total.

Table 2.3 shows how many responses were manually assigned to the several categories of motivations. The majority of the respondents appear to be concerned about their privacy (61.3%), while 13.1% does not give a specific reason.

Table 2.3 Category prevalence of the manually coded responses in LISS panel (multi-label).

Category	LISS (manual) N	%
Privacy	1115	61.3
Misuse	139	7.6
Safety	113	6.2
Effort	100	5.5
Control	187	10.3
Surveillance	100	5.5
Elsewhere	12	0.7
Without reason	239	13.1
Due to emotions	87	4.8
Other/don't know	22	1.2
Non-informative	61	3.4
Total	2175	100

3 Methods

The data of the consent survey are analyzed by automatically classifying the open answers into categories. For this purpose, we assume the same categories of motivations that are used for the LISS panel. After cleaning texts for both sources, we extracted salient keywords with NLP and a rule-based language model, and then estimated category probabilities with penalized logistic regression. We trained and evaluated on the manually coded LISS data and then apply the models to the consent survey. Analyses were run in RStudio (2024.04.1).

3.1 Text preprocessing

Data cleaning

The first step is to clean the data. As mentioned in Subsection 2.5, not all responses were in Dutch. Non-Dutch text fields were translated into Dutch, and as many spelling errors as possible were corrected. We used the R package `textcat` (Hornik et al. 2023) for language detection, DeepL Translator for translation, and Microsoft Word's spell checker for spelling correction. In addition, digits, punctuation, and double spaces were removed, and all uppercase letters were converted to lowercase. Single dots were retained to preserve sentence boundaries. The package `textcat` detected 6 English records for the LISS panel study and 1 for the consent survey.

Because punctuation is removed during cleaning, responses consisting only of punctuation marks are dropped. For the LISS panel study, this concerns 10 responses in the non-informative category, leaving 1,808 responses in the LISS training set. Table 3.1 reports how many responses remain per question after cleaning. In our analyses, survey responses consisting only of punctuation are treated as non-informative.

Table 3.1 Frequency of responses for the consent survey before (raw) and after cleaning (clean).

Survey	Question	Raw responses	Clean responses
LISS panel study	Total	1818	1808
Statistics	Selfie	1399	1387
Netherlands	Location	554	554
Consent Survey	House	1505	1503
	Receipt	1355	1348
	Video	1459	1457
	Final remarks	865	843
	Total	7137	7092

3.2 Keyword extraction

We then distilled the cleaned texts for both the LISS panel study and the consent survey by extracting the most informative word sequences. Using NLP, we selected one or more word combinations per sentence that capture its essence. For example, the sentence "no assurance can be given on the collection of my private data." is summarised with the keywords "no assurance", "collection", and "private data".

Sentences are linguistically decomposed into tokens (*tokenisation*), parts of speech (*POS tagging*), and dependency relations (*dependency parsing*). We use the R package `udpipe` (Wijffels 2023; Straka et al. 2016), which relies on a deep-learning language model. Table 3.2 shows sample output for the example sentence. This sentence has 13 tokens; each token points to a parent token. For both tokens and parents, the universal part-of-speech (upos) tags are identified and for every token the universal dependency relationship (dep_rel) with its parent is determined. For an overview of universal part-of-speech tags and types of universal dependency relationships in different languages, see Universaldependencies.org (2014-2024). For each token and parent, the stem and the lemma were also determined.

Table 3.2 NLP results for the sentence "no assurance can be given on the collection of my private data." where `upos` displays universal part-of-speech tags and `dep_rel` the dependency relationships between tokens and parents.

term_id	token	upos token	parent	upos parent	dep_rel
1	no	determiner	assurance	noun	determiner
2	assurance	noun	given	verb	passive nominal object
3	can	auxiliary	given	verb	auxiliary
4	be	auxiliary	given	verb	passive auxiliary
5	given	verb	<NA>	<NA>	root
6	on	adposition	collection	noun	case marking
7	the	determiner	collection	noun	determiner
8	collection	noun	given	verb	oblique nominal
9	of	adposition	data	noun	case marking
10	my	pronoun	data	noun	possessive nominal modifier
11	private	adjective	data	noun	adjectival modifier
12	data	noun	collection	noun	nominal modifier
13	.	punctuation	given	verb	punctuation

Based on the linguistically decomposed sentences, the most relevant keywords were selected by means of a rule-based language model, focusing on word combinations with an adjective, noun, or verb as parent. The number of selected keywords per sentence varies and is at least one. Each

keyword may consist of one term, a bigram, or a trigram. Thus, when an answer consists of three sentences, at least three keywords are selected. The language model is given in Appendix B. For the example above, the model returns “no assurance” (step 2a), “private data” (step 2b), and “collection” (step 2c).

Table 3.3 lists, for each question, the total number of keywords extracted and the five most frequent keywords. As expected for this domain, “privacy” and “private” are very common.

Table 3.3 Frequency of keywords per question

Survey	Question	Keywords	Most frequent keywords
LISS panel study	Why not share?	1262	“privacy”, “private”, “data”, “photo”, “personally”
Statistics Netherlands consent survey	Selfie	553	“privacy”, “private”, “photo”, “not necessary”, “survey”
	Location	378	“privacy”, “location”, “not necessary”, “private”, “survey”
	House	644	“privacy”, “private”, “house”, “photo”, “don’t want”
	Receipt	558	“privacy”, “no receipt”, “private”, “receipt”, “don’t see”
	Video	643	“privacy”, “private”, “video”, “surroundings”, “not necessary”
	Final remarks	786	“no”, “questions”, “questionnaire”, “none”, “personally”

3.3 Classification model

Classification of the open-ended responses is performed by predicting probabilities for the motivation categories using logistic regression. We assume a logistic relationship between a response’s keywords and its probability for a given category:

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta' x_i + e_i, \text{ with } e_i \sim \mathcal{N}(0, \sigma^2). \tag{1}$$

Here, p_i is the probability that response i belongs to the given category, x_i is a q-vector containing the frequencies of the keywords in response i, β' is the q-vector of regression coefficients, and e_i a normally distributed error term.

Before applying model (1) to the consent-survey responses, we fit it to the manually coded LISS responses. We estimate an independent model for each of the 11 categories. Ten percent of LISS responses are held out as test data; the remaining 90% are used for fitting. The test set is constructed by randomly selecting 10% within each category.

The regression coefficients β in model (1) are estimated by minimizing

$$\sum_{i=1}^{n} \left(\text{logit}(p_i) - \beta' x_i\right)^2 + \lambda \sum_{j=1}^{q} \left[\frac{1}{2}(1-\alpha)\beta_j^2 + \alpha|\beta_j|\right]. \tag{2}$$

Here, n is the number of responses in the given category of the training set and q is the dimension of x_i and β, so q corresponds to the total number of keywords in the training set of a given category. The parameter $\lambda \geq 0$ is a penalty factor on the regression coefficients. If $\lambda = 0$, the penalized regression coincides with the usual (logistic) regression. Further, $0 \leq \alpha \leq 1$ indicates whether the penalty is applied to the absolute values or the squares of the regression coefficients. For $\alpha = 0$ (ridge), the penalty is only on the squares and works well if there are multiple predictors with roughly equal effects. For $\alpha = 1$ (lasso), the penalty is applied to the absolute values and works well if there are relatively few predictors with larger effects.

3.4 Training and validation

Optimization of the parameters α and λ is done through a 10-fold cross-validation with AUC (Area Under the ROC Curve) as quality measure. The AUC is defined as the area under the ROC (Receiver Operating Characteristic) curve (Swets 1996; Fawcett 2006) and can be interpreted as the probability that a randomly selected response from a given category gets a higher p_i value than a randomly selected response from another category. The purpose is to maximize this probability. Cross-validation is a technique to determine the predictive power of a model (Allen 1974; Stone 1977) and is frequently used in text mining to fit models (Seni and Elder 2010). We used the R-package `glmnet` (Friedman et al. 2010; Tay et al. 2023) for this purpose. Once the regression coefficients β are estimated for a given category based on the LISS data, model (1) is used to predict the probability p_i that a response i of the consent survey belongs to the given category.

Because AUC assesses ranking performance independent of any threshold, we additionally report thresholded classification metrics that evaluate the quality of the final category assignments at a prespecified decision threshold $d = 0.2$ (see also Section 4.1). Specifically, we compute per-class precision, recall, and F_1, as well as macro-averaged and micro-averaged F_1. These measures complement AUC by quantifying the trade-off between false positives and false negatives at the chosen threshold, which is useful under class imbalance and in multi-label settings. These metrics are derived using the manual labels available for validation.

4 Results

We first motivate the decision threshold d for multi-label assignment (Section 4.1) and present model validation on the *LISS panel study*. We then answer the research questions: overall prevalence of motivation categories across both questionnaires (RQ1; Section 4.4), per-sensor distributions (RQ2; Section 4.5), and heterogeneity by background characteristics (RQ3; Section 4.6).

4.1 Threshold choice and multi-label incidence

For the *Statistics Netherlands consent survey*, category assignment is automated using a probability threshold d. Table 4.1 reports how many categories per answer are assigned at different d. As a compromise between coverage and avoiding many multi-assignments, we set $d = 0.2$ for all downstream analyses: only 0.5% of answers remain unassigned, while 71.8% receive exactly one label.

Table 4.1 Consent survey: number (percentage) of automatic assignments across thresholds d, by number of categories per answer ($n = 7{,}137$).

d	Number of categories 0 n (%)	1 n (%)	2 n (%)	3 n (%)	4 n (%)	5 n (%)	6 n (%)
0.1	1 (0.0)	4101 (57.5)	2147 (30.1)	746 (10.5)	115 (1.6)	18 (0.3)	9 (0.1)
0.2	39 (0.5)	5122 (71.8)	1801 (25.2)	156 (2.2)	19 (0.3)	0 (0.0)	0 (0.0)
0.3	182 (2.6)	6199 (86.9)	706 (9.9)	42 (0.6)	8 (0.1)	0 (0.0)	0 (0.0)
0.4	2752 (38.6)	4194 (58.8)	171 (2.4)	19 (0.3)	1 (0.0)	0 (0.0)	0 (0.0)
0.5	2983 (41.8)	4024 (56.4)	118 (1.7)	11 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)

4.2 Model validation on LISS

Table 4.2 reports cross-validated (α, λ) and test-set AUC per category for the LISS models; Table 4.3 lists the most predictive keywords (positive/negative). *Privacy*, *Safety*, and *Due to emotions* achieve very high AUC (> 0.90), whereas *Elsewhere* and *Other/don't know* are harder to predict (few cases).

Table 4.2 Cross-validation results for LISS categories: optimal α and λ, AUC in the test set, and number of predictive keywords with positive/negative effects.

Category	α	λ	AUC	positive keywords	negative keywords
Privacy	0.2	0.032	0.933	199	336
Misuse	0.1	0.031	0.885	180	57
Safety	0.2	0.048	0.971	140	5
Effort	0.1	0.064	0.825	111	18
Control	0.1	0.079	0.821	221	16
Surveillance	0.4	0.013	0.716	106	30
Elsewhere	0.1	<0.01	0.678	27	64
Without reason	0.6	<0.01	0.869	165	453
Due to emotions	0.3	<0.01	0.996	89	59
Other/don't know	0.1	<0.01	0.430	48	24
Non-informative	0.3	<0.01	0.876	24	95

Furthermore Table 4.4 shows the thresholded per-class precision, recall, and F_1 on the LISS panel using out-of-fold predictions at $d = 0.2$, reporting macro- and micro-averaged F_1. The thresholded metrics indicate excellent in-domain performance for clearly delineated categories such as *Due to emotions* and *Non-informative* (both $F_1 \geq 0.93$), and strong results for *Misuse*, *Safety* and *Without reason* ($F_1 \approx 0.87$–0.90). Categories with more diffuse semantic boundaries (*Surveillance*, *Other/don't know*) perform somewhat lower but remain solid ($F_1 \approx 0.80$–0.84). The micro-averaged $F_1 = 0.95$ reflects high overall assignment quality on LISS.

Table 4.3 Predictive keywords for categories in the LISS panel study.

Category	Sign	Keywords with biggest positive or negative effect
Privacy	+	“due to privacy”, “no privacy”, “personally”, “private”, “privacy”
	—	“google facebook”, “system”, “don’t send”, “no”, “gps”
Misuse	+	“personal life”, “abuse fraud”, “name”, “insecurity”, “commercial company”
	—	“privacy violation”, “products”, “quantity”, “anonymous”, “big”
Safety	+	“other hands”, “wrong purposes”, “safe”, “no certainty”, “wrong hands”
	—	“lower”, “clients”, “stories”, “research purposes”, “threshold”
Effort	+	“visibility”, “work”, “no need”, “work location”, “personal effort”
	—	“find unpleasant”, “daily life”, “desks”, “university point of view”, “privacy”
Control	+	“degree”, “scary”, “not obviously”, “no importance”, “no use”
	—	“internet”, “home”, “environment”, “meaningful”, “privacy”, “well-known”
Surveillance	+	“not search”, “insight”, “google facebook”, “brother”, “big brother”
	—	“need”, “nonsense”, “case”, “bank details”, “known opinion”
Elsewhere	+	“other possibilities”, “nonsense”, “knows”, “other sources”, “info”
	—	“vaguely”, “real”, “effort”, “commercial purposes”, “registered”
Without reason	+	“no video”, “recordings not possible”, “no images”, “choose not”, “daily life”
	—	“written questions”, “privacy”, “personal”, “major attack”, “invasion”
Due to emotions	+	“personal feeling”, “unsafe”, “uncomfortable”, “not pleasant”
	—	“worth”, “etc”, “photos videos”, “unknown people”, “private situation”
Other/don’t know	+	“limitation”, “affected”, “incentive”, “reflect”, “no social media”
	—	“more cumbersome”, “such information”, “humanity”, “sorry”, “usefulness”
Non-informative	+	“not”, “no comment”, “n a”, “na”, “no”
	—	“no selfies”, “night”, “marketing”, “shower”, “cumbersome”

Table 4.4 Per-class precision, recall, and F_1 for the LISS panel at threshold $d = 0.2$ (out-of-fold predictions).

Category	Precision	Recall	F_1
Privacy	0.70	1.00	0.83
Misuse	0.92	0.88	0.90
Safety	0.87	0.86	0.87
Effort	0.79	0.84	0.81
Control	0.77	0.86	0.82
Surveillance	0.80	0.80	0.80
Elsewhere	0.92	0.92	0.92
Without reason	0.82	0.93	0.87
Due to emotions	0.91	0.99	0.95
Other/don’t know	0.86	0.82	0.84
Non-informative	0.91	0.96	0.93
Macro average	0.84	0.90	0.87
Micro average	0.95	0.95	0.95

4.3 Qualitative error analysis

From Table 4.5 it follows that 39 consent-survey answers remain unassigned at $d = 0.2$. Many are short answers using weakly predictive words (e.g., “I don’t do it”), especially for *Selfie*. Some vocabulary is highly specific (e.g., “no obligation need”, “unfunny joke”) or survey-specific (UI remarks in final comments). We manually reviewed and coded these 39 answers to maintain full coverage.

A handful of answers labeled *Other/don’t know* suggest two survey-specific reasons in the consent survey: (i) technical problems in uploading, GPS activation, or media capture; and (ii)

confusion about instructions (e.g., inside vs. outside house). These did not warrant extending the taxonomy here but can inform future instrument refinements.

Table 4.5 Automatic assignments per consent-survey question at $d = 0.2$, by number of categories.

		Number of categories				
Question	Responses	0	1	2	3	4
Selfie	1399	17 (1.2%)	1033 (73.8%)	319 (22.8%)	27 (1.9%)	3 (0.2%)
Location	554	1 (0.2%)	349 (63.0%)	182 (32.9%)	21 (3.8%)	1 (0.2%)
House	1505	7 (0.5%)	1057 (70.2%)	408 (27.1%)	29 (1.9%)	4 (0.3%)
Receipt	1355	8 (0.6%)	1080 (79.7%)	255 (18.8%)	11 (0.8%)	1 (0.1%)
Video	1459	2 (0.1%)	958 (65.7%)	446 (30.6%)	50 (3.4%)	3 (0.2%)
Final remarks	865	4 (0.5%)	645 (74.6%)	191 (22.1%)	18 (2.1%)	7 (0.8%)
Total	7137	39 (0.5%)	5122 (71.8%)	1801 (25.2%)	156 (2.2%)	19 (0.3%)

Note: No responses were assigned to 5 or 6 different categories for any task.

4.4 Overall prevalence of motivation categories (RQ1)

Table 4.6 summarizes how often each motivation category occurs, side by side for both surveys and methods. For the LISS panel, manual counts (left block) sum to $n = 2{,}175$ because an answer may receive multiple categories (multi-label). For the consent survey, the right block pools across all five sensor tasks and the final-remarks item ($n = 7{,}137$ answers), again allowing multiple categories per answer. Percentages use the number of answers as denominator.

Across both sources, *Privacy* dominates, with high prevalence in the consent survey pool (86%) and the LISS training labels (61%). *Control* and *Without reason* are the next most common. Rare categories such as *Elsewhere* and *Other/don't know* remain infrequent.

Table 4.6 Overall category prevalence by survey and method (multi-label).

	LISS (manual)		LISS (model)		Consent (model)	
Category	N	%	N	%	N	%
Privacy	1115	61.3	1583	58.5	6157	86.3
Misuse	139	7.6	133	4.9	83	1.2
Safety	113	6.2	111	4.1	36	0.5
Effort	100	5.5	107	4.0	179	2.5
Control	187	10.3	208	7.6	727	10.2
Surveillance	100	5.5	100	3.7	101	1.4
Elsewhere	12	0.7	12	0.4	47	0.7
Without reason	239	13.1	272	10.1	1174	16.5
Due to emotions	87	4.8	94	3.5	122	1.7
Other/don't know	22	1.2	21	0.8	41	0.6
Non-informative	61	3.4	64	2.4	640	9.0
Total	2175	100	2705	100	7137	100

4.5 Category distributions by task and final remarks (RQ2)

We next quantify how often each motivation category occurs by sensor task and for the *final remarks* in the consent survey. Figures 4.1-4.3 and Table 4.7 summarize the patterns.

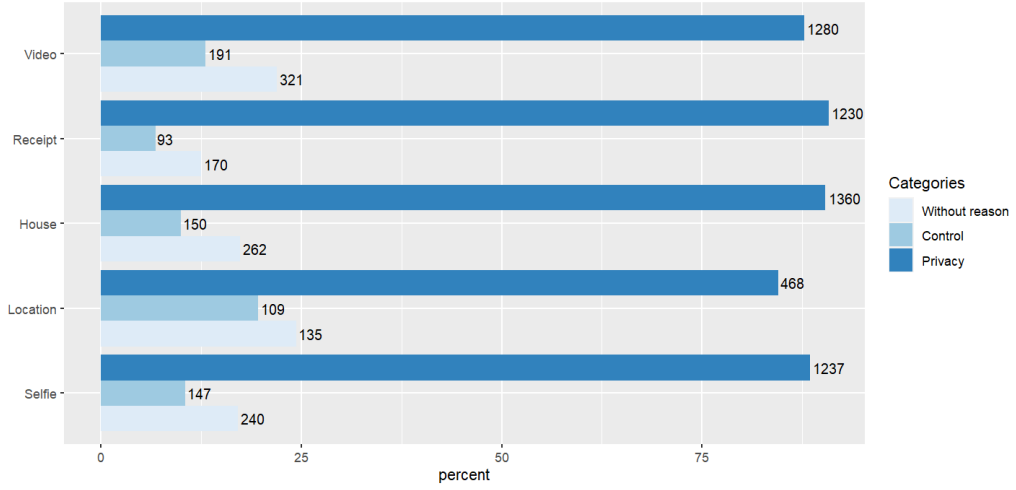
Table 4.7 Frequency of classified responses by sensor task (consent survey).

Category	Task Selfie	Location	House	Receipt	Video
Privacy	1237 (88.4%)	468 (84.5%)	1360 (90.4%)	1230 (90.8%)	1280 (87.7%)
Misuse	7 (0.5%)	12 (2.2%)	6 (0.4%)	7 (0.5%)	2 (0.1%)
Safety	5 (0.4%)	8 (1.4%)	4 (0.3%)	5 (0.4%)	9 (0.6%)
Effort	20 (1.4%)	6 (1.1%)	55 (3.7%)	33 (2.4%)	44 (3.0%)
Control	147 (10.5%)	109 (19.7%)	150 (10.0%)	93 (6.9%)	191 (13.1%)
Surveillance	16 (1.1%)	12 (2.2%)	21 (1.4%)	8 (0.6%)	21 (1.4%)
Elsewhere	10 (0.7%)	3 (0.5%)	10 (0.7%)	10 (0.7%)	11 (0.8%)
Without reason	240 (17.2%)	135 (24.4%)	262 (17.4%)	170 (12.5%)	321 (22.0%)
Due to emotions	38 (2.7%)	14 (2.5%)	13 (0.9%)	10 (0.7%)	34 (2.3%)
Other/don't know	6 (0.4%)	3 (0.5%)	14 (0.9%)	7 (0.5%)	8 (0.5%)
Non-informative	55 (3.9%)	11 (2.0%)	88 (5.9%)	62 (4.6%)	93 (6.4%)
Total	1399	554	1505	1355	1459

Privacy is by far the most frequently assigned category for every task. The prevalence is lowest for *Location* (84.5%) and highest for the camera-based tasks *Receipt* (90.8%) and *House* (90.4%), with *Selfie* and *Video* in between (88.4% and 87.7%, respectively). These levels are consistent with the centrality of privacy concerns observed in LISS and confirm that, regardless of modality, respondents primarily frame their reluctance in terms of privacy.

Beyond privacy, *Control* is the most common secondary reason and is particularly salient for *Location* (19.7%). Open answers often state that the precise location is not relevant to the survey or express a desire to retain agency over what is shared ("my exact location is not relevant for this survey"; "I do not see the added value"). *Video* also shows a relatively high incidence of *Control* (13.1%), followed by *Selfie* (10.5%) and *House* (10.0%); *Receipt* is lowest (6.9%).

Figure 4.1 Classified responses by sensor task: Privacy, Control, Without reason (consent survey).



In parallel, *Surveillance* concerns cluster around *Location* (2.2%) and, to a lesser extent, *House* and *Video* (both 1.4%), consistent with remarks about traceability ("it makes me traceable") and a general unease with being tracked. Respondents seem to have the most difficulty with taking a photo of their house (3.7%), because it requires people to go outside: "I don't feel like going into the street." or "I'm still in bed and not dressed.". Sharing a selfie, video or location clearly evokes more emotions than sharing a photo of the house or receipt.

Effort is most frequently mentioned for *House* (3.7%), followed by *Video* (3.0%) and *Receipt* (2.4%); it is least common for *Location* (1.1%). Qualitative comments for *House* often point to the practical inconvenience of going outside to capture a suitable photo (e.g., being indoors, weather, time of day). *Due to emotions* (e.g., "uncomfortable", "not a pleasant idea") is more often linked to *Selfie* (2.7%), *Video* (2.3%), and *Location* (2.5%) than to *House* (0.9%) or *Receipt* (0.7%), suggesting a stronger affective component when sharing personally revealing imagery or live context.

Figure 4.2 Classified responses by sensor task: Effort, Surveillance, Due to emotions, Non-informative (consent survey).

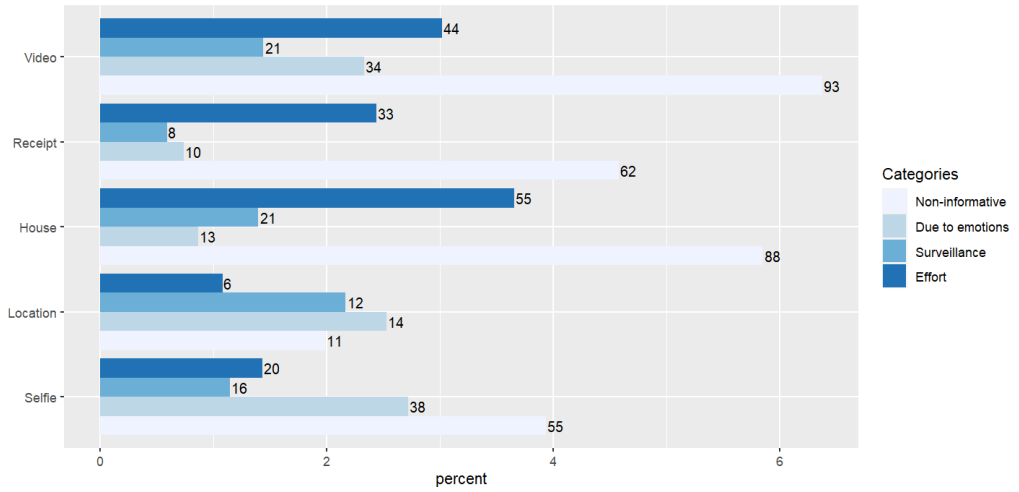
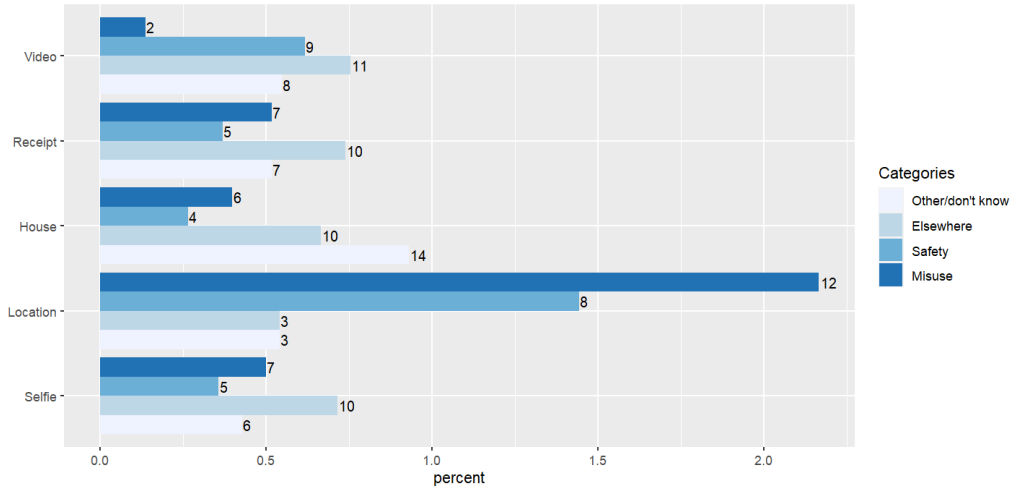


Figure 4.3 Classified responses by sensor task: Misuse, Safety, Elsewhere, Other/don't know (consent survey).



Mentions of *Misuse* and *Safety* are rare across all tasks. Where present, they appear slightly more often for *Location* (Misuse 2.2%; Safety 1.4%) than for the camera-based tasks (generally $\leq$0.6%). *Elsewhere* and *Other/don't know* are similarly infrequent (each $\leq$0.9% per task). *Non-informative* answers (e.g., "no", "n/a") remain limited in the by-task view (about 2–6%), highest for *House* (5.9%) and *Video* (6.4%).

The category *Without reason* captures brief refusals without a substantive motivation. It is most prevalent for *Video* (22.0%) and *Location* (24.4%), and lower for *Receipt* (12.5%). This pattern aligns with the idea that some tasks feel optional or not obviously relevant to the survey topic for

some respondents, while others (e.g., a receipt) may be refused more deliberately with an accompanying rationale.

For the *final remarks* (Table 4.8), *Privacy* again dominates (67.3%), while *Non-informative* is comparatively high (38.3%) because many respondents explicitly write that they have no additional comments ("none", "nothing"). Substantive categories appear with low frequency; among those, *Control* (4.3%) and *Misuse* (5.7%) are more common than, for example, *Safety* (0.6%) or *Elsewhere* (0.3%). Illustrative *Effort* remarks in the final box concern usability (slow loading, small screen, layout), underlining that a fraction of refusal-related comments reflect practical barriers rather than strictly attitudinal ones. For example: "I filled out the questionnaire on my iphone and took quite a long time to load and I had to shift the image to read the whole sentences." and "Let a good web developer build this kind of thing. People are going to drop out if the questions are so hard to read." Emotions that are mentioned in the category due to emotions are uncomfortable, unpleasant and also scary.

Table 4.8 Frequency of classified final remarks (consent survey).

Category	Frequency (percent)
Privacy	582 (67.3%)
Misuse	49 (5.7%)
Safety	5 (0.6%)
Effort	21 (2.4%)
Control	37 (4.3%)
Surveillance	23 (2.7%)
Elsewhere	3 (0.3%)
Without reason	46 (5.3%)
Due to emotions	13 (1.5%)
Other/don't know	3 (0.3%)
Non-informative	331 (38.3%)
Total	865

A few responses were classified to the category other/don't know, both for the final remarks and for the different types of sharing. Based on these responses, we could identify two new categories for not wanting to share information in the consent survey. The first category is that respondents encounter technical difficulties in sharing information: "I can't take a picture because uploading the picture didn't work.", "uploading photo does not work", "the gps option did not work.". The second category is that respondents do not fully understand the questionnaire: "no idea what exactly the question is photo of inside or outside house.", "instructions are not clear.".

4.6 Heterogeneity by demographics (RQ3)

We examine whether the observed reasons in the consent survey differ by gender, age, and education. For final remarks (Figures 4.4–4.6), we find no notable gender differences in *Privacy*, *Misuse*, *Effort*, and *Non-informative*. Men more often mention *Control* (56.8%), whereas women score higher on *Due to emotions* (77.0%), *Surveillance* (73.9%), and *Without reason* (71.7%). Small categories (e.g., *Elsewhere*, *Safety*) have too few cases for stable comparisons.

By age, *Privacy* and *Non-informative* are flat. Younger groups mention *Misuse*, *Surveillance*, and *Control* more than older groups; *Without reason* is highest at ages 16–29 (43.5%) and lowest at 68+ (2.2%). *Effort* is highest at ages 50–67 (42.9%).

By education, differences are small overall, with somewhat higher rates among higher-educated for *Surveillance* (52.2%) and *Effort* (55.0%).

Figure 4.4 Final-remarks classifications by gender.

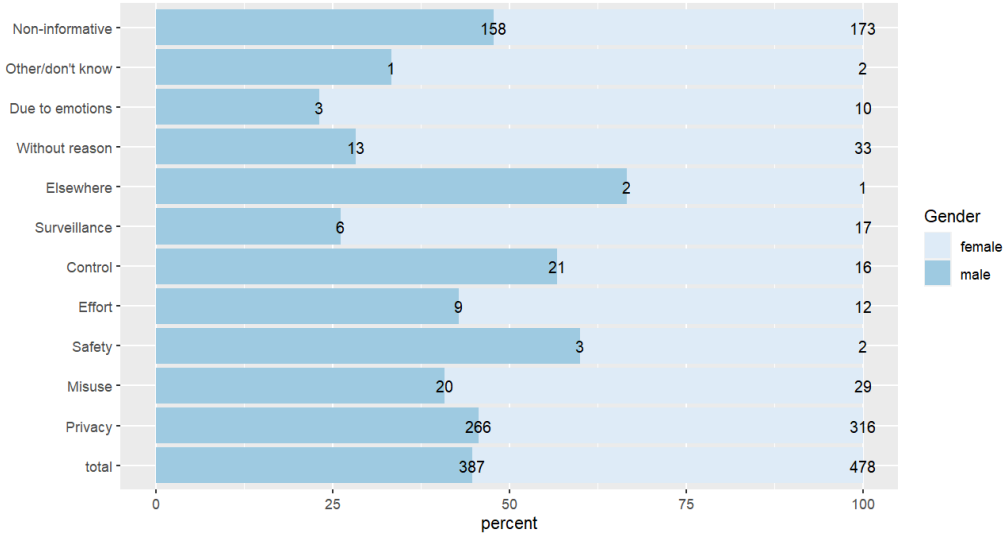
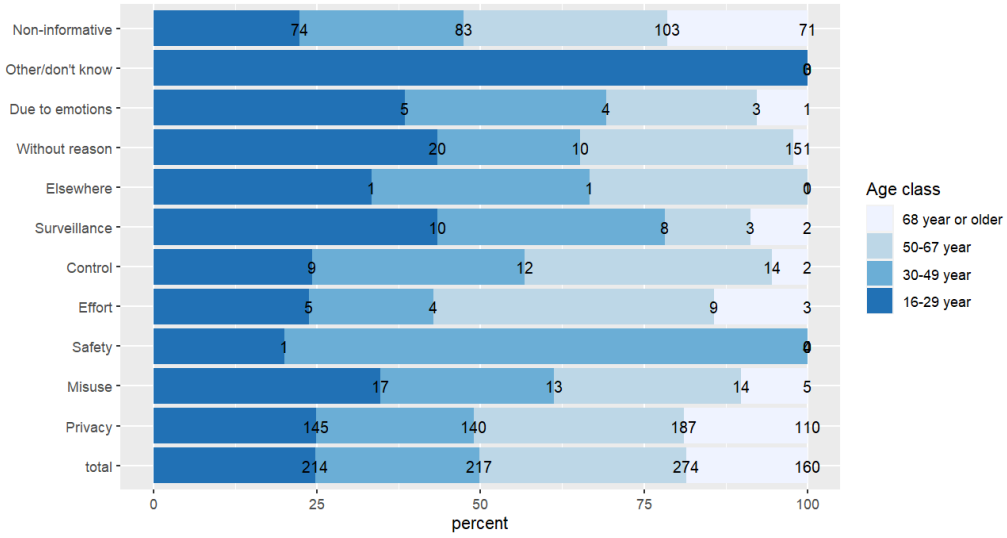


Figure 4.5 Final-remarks classifications by age class.



We then restrict to the three most prevalent categories (*Privacy*, *Control*, *Effort*) and study per-sensor heterogeneity (Figures 4.7–4.15). For *Privacy*, we see no major demographic differences. For *Control*, women score higher than men for all tasks except *Location*. By age, *Control* is highest at 30–49 for *Location*, *Video*, and *Selfie*, but older groups score higher on *Receipt*. Higher-educated respondents show higher *Control* for *Selfie*, *House*, and *Location*.

For *Effort*, women more often report effort for *Video* and *House*, men for *Receipt*. Ages 50–67 report more effort for *Video*, *Selfie*, and *Receipt*; ages 16–29 report more difficulty with *House*.

Figure 4.6 Final-remarks classifications by education.

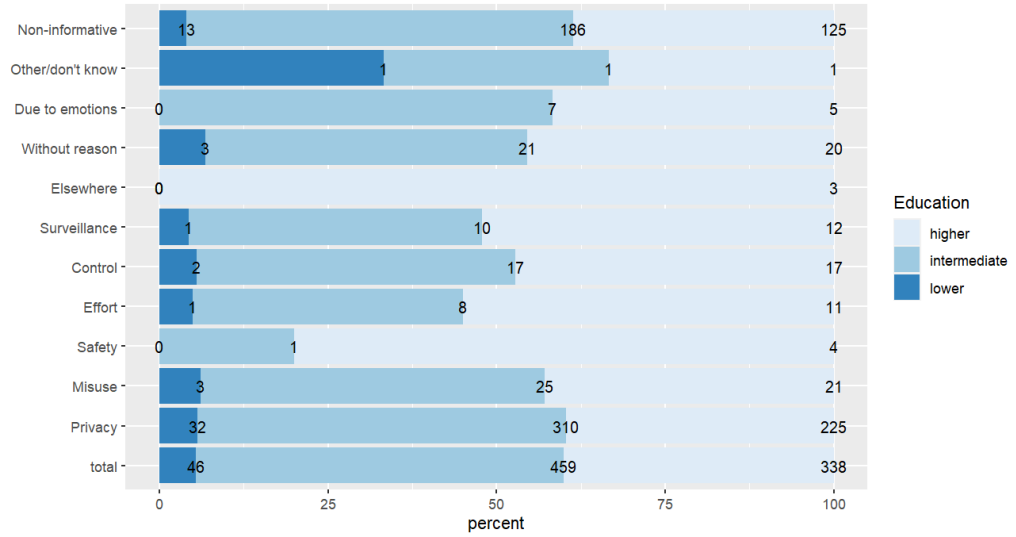


Figure 4.7 Per-sensor distributions by gender: total and Privacy.

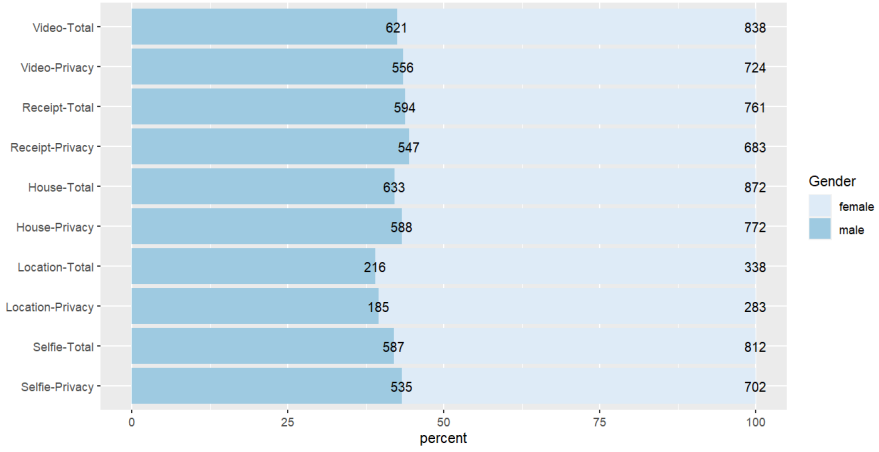


Figure 4.8 Per-sensor distributions by age: total and Privacy.

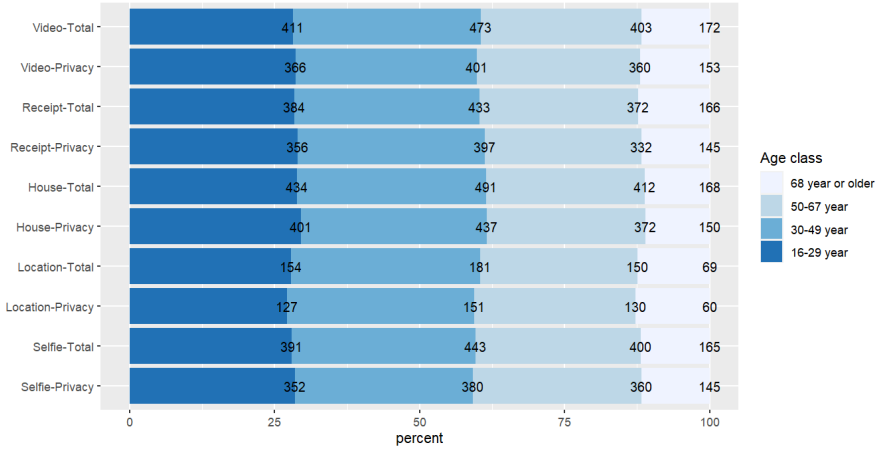


Figure 4.9 Per-sensor distributions by education: total and Privacy.

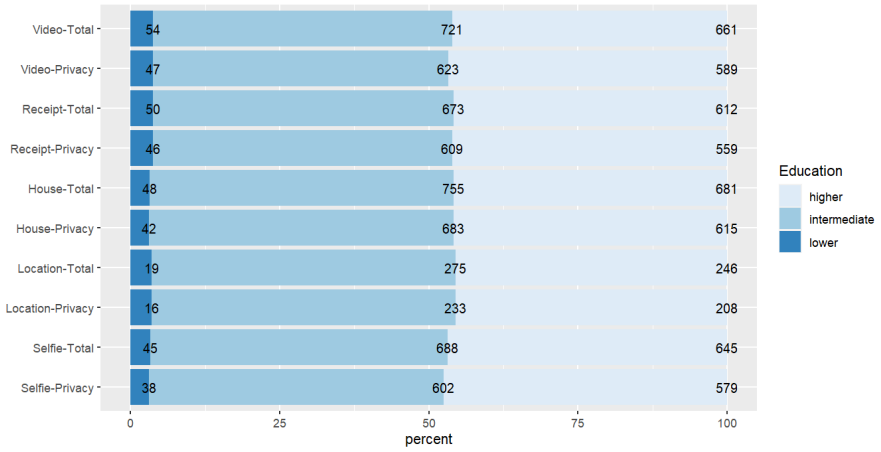


Figure 4.10 Per-sensor distributions by gender: total and Control.

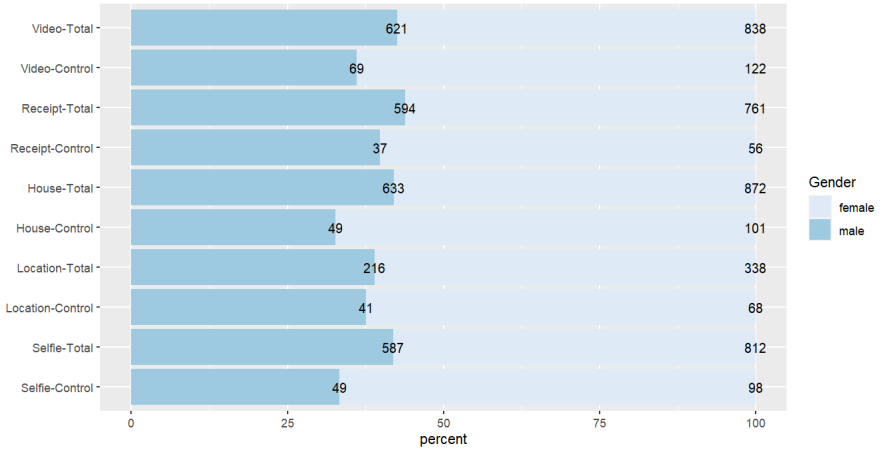


Figure 4.11 Per-sensor distributions by age: total and Control.

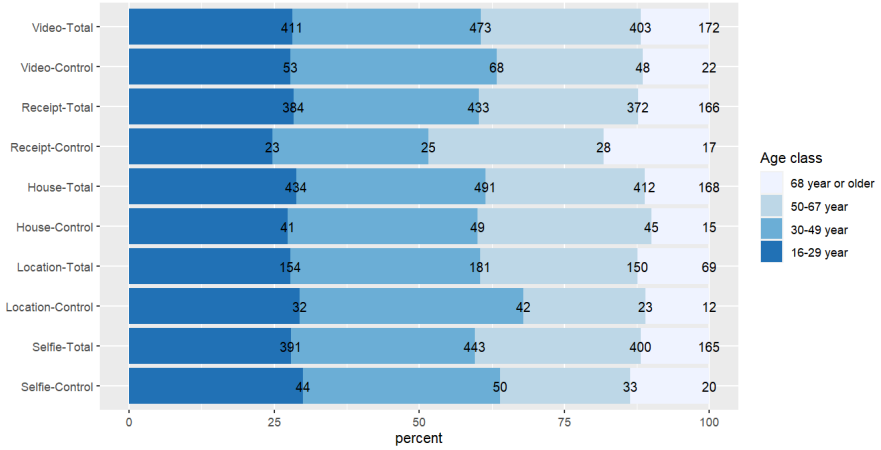


Figure 4.12 Per-sensor distributions by education: total and Control.

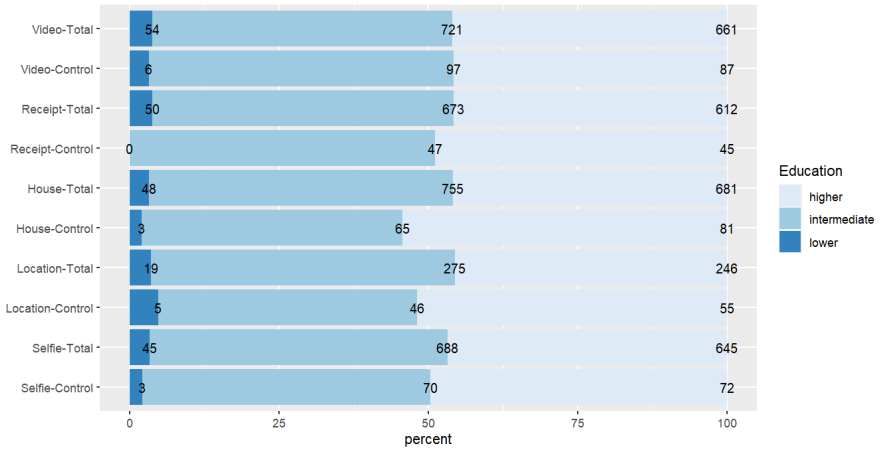


Figure 4.13 Per-sensor distributions by gender: total and Effort.

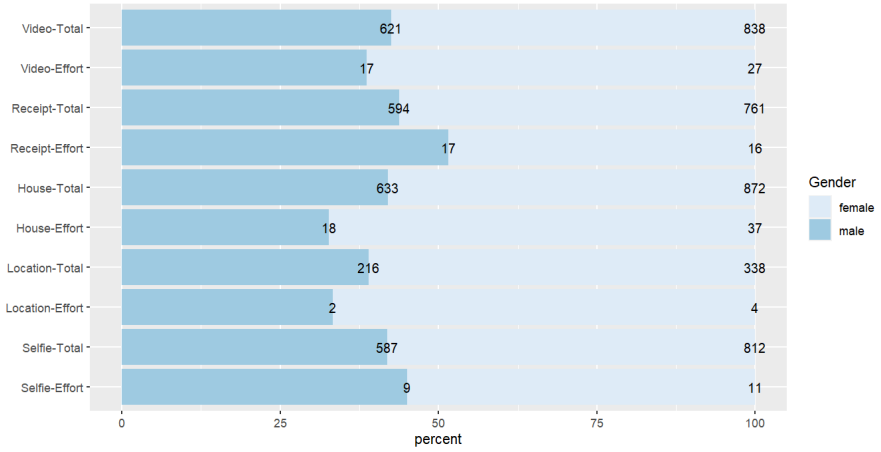


Figure 4.14 Per-sensor distributions by age: total and Effort.

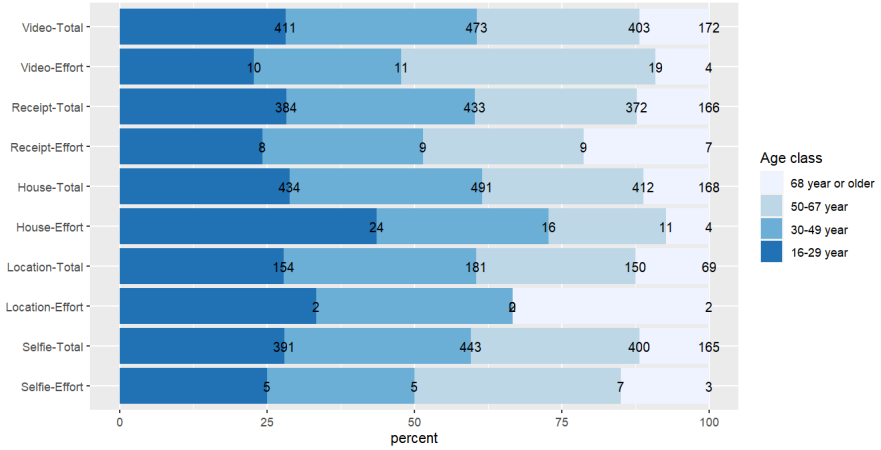
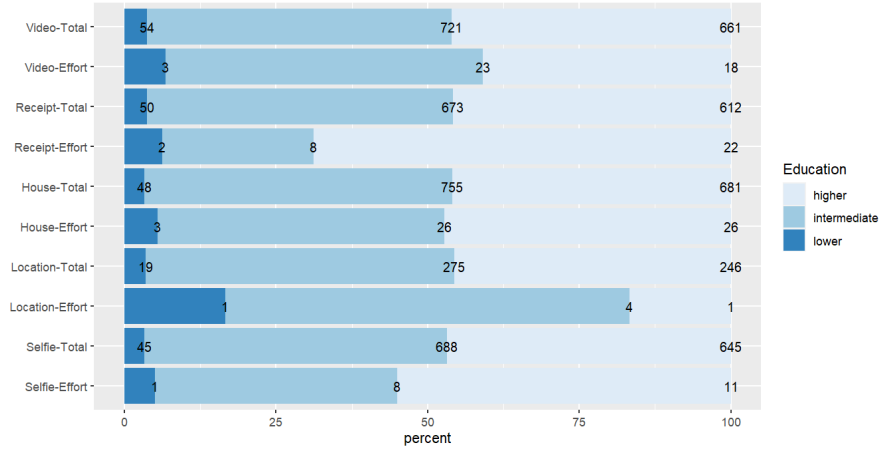


Figure 4.15 Per-sensor distributions by education: total and Effort.



5 Discussion

This study set out to classify short Dutch open-ended answers about (un)willingness to share smartphone-sensor data using an interpretable pipeline that transfers an existing motivation taxonomy from the LISS panel study to the Statistics Netherlands consent survey.

The overall results across questionnaires are consistent and intuitive. Privacy dominates as the primary motivation, both in the pooled statistics and within each sensor task . In the consent survey, privacy is also the leading category in the final-remarks item, with a sizeable share of non-informative entries, unsurprising given the lower response fraction to that optional box. While privacy concerns are widely shared, the analyses reveal task-specific accents: location sharing elicits comparatively more control and surveillance concerns, whereas the house-photo task shows the strongest effort signals, plausibly because it may require going outside or adjusting one's environment. Emotions (e.g., feeling uncomfortable) feature more often for selfie, video, and location than for house or receipt. Instances labeled other/don't know surfaced two practical themes unique to the consent survey context: technical difficulties (e.g., upload failures, GPS not working) and comprehension issues about instructions. These suggest concrete opportunities for instrument refinement.

Subgroup patterns add nuance. For the final-remarks item, women more often mention emotions and surveillance while men mention control, yet the per-task breakdown suggests women indicate control more frequently for most camera-based tasks; taken together, these results underline that patterns at the instrument level need not mirror patterns at the task level. Younger respondents report surveillance, control, and misuse more often, whereas effort is more frequently cited by respondents aged 50–67. Differences by education are modest; where present, higher-educated respondents more often register control for selfie, house, and location, and report somewhat more effort for selfie, receipt, and house. Because several categories are rare and splits can be sparse, such subgroup contrasts should be interpreted cautiously.

The findings point to three immediate implications for future smartphone-sensor questionnaires in official statistics. First, because privacy is the modal concern across all tasks, communications should address it explicitly and concretely: what is collected, why it is needed, how long it is kept, who may access it, and how respondents can exercise control or revoke permission. The emphasis should be strongest for location, where surveillance and control concerns are especially salient. Second, the effort signal around the house-photo task suggests small design

adjustments could reduce burden, clearer prompts about framing and what to include or exclude, acceptance of existing photos where feasible, and an obvious “skip” option. Third, the two survey-specific signals highlight the value of brief on-screen troubleshooting and tighter wording; an illustrative mock-screen or example image before opening the camera or map could reduce confusion and break-offs. These are low-cost changes that target precisely the pain points observed in the open answers.

Methodologically, the combination of rule-based keyword extraction and penalized logistic models strikes a useful balance between transparency and performance for short survey texts. The threshold analysis indicates that a decision threshold of $d = 0.2$ yields a practical compromise: very few answers remain unassigned while multi-label inflation is limited (Table 2.2). Crucially, substantive patterns by task and subgroup are stable when d varies within a reasonable range, which supports robustness. Cross-validated AUCs on the LISS panel (Table 4.2) are high for the most substantive categories (e.g., Privacy, Safety, Due to emotions) and lower for diffuse or rare categories (Elsewhere, Other/don’t know), aligning with coder experience that these latter boundaries are inherently fuzzy. The most predictive keywords per category (Table 4.3) are face-valid and interpretable, a feature that eases communication with domain experts and facilitates targeted instrument changes.

From a validation perspective, it is helpful to distinguish threshold-agnostic discrimination from thresholded assignment quality. AUC summarizes ranking performance and guided hyperparameter tuning, while per-class precision, recall, and F_1 at $d = 0.2$ quantify the trade-off between false positives and false negatives in the final classifications. On the LISS panel, these thresholded metrics complement AUC and confirm that strongly delineated categories perform best.

Several limitations deserve mention. Manual labels exist for the LISS panel but not for the Statistics Netherlands consent survey; as a result, we treat the transferability of the LISS-trained classifiers to the consent survey answers as a working assumption rather than as a validated fact. Both surveys are closely related, the questionnaires are near-identical, and the studies were conducted in quick succession. Also, the strong face-validity of patterns and the agreement in aggregate prevalence across questionnaires make this assumption plausible, but they do not constitute hard evidence. To really confirm out-of-domain performance, a calibration set for the consent survey would be needed.

In sum, a transparent NLP pipeline leveraging an existing taxonomy transfers well from the LISS panel study to the consent survey and yields coherent, actionable insights. Privacy concerns are pervasive, control and surveillance concentrate around location, and effort is most pronounced for the house-photo task; these patterns persist across reasonable threshold choices and are consistent with prior work on willingness to share sensor data. The approach scales to large volumes of short survey texts while remaining interpretable, which is important for governance and for closing the loop between classification results and questionnaire design. Establishing and maintaining a small gold-label set will make this pipeline a practical, repeatable component of survey production, enabling statisticians to monitor concerns over time and to adapt questionnaires and communication materials in a targeted, evidence-based manner.

Acknowledgments

The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands. The authors are grateful to Chris Lam for his comments.

References

Allen, D. (1974). “The Relationship between Variable Selection and Data Agumentation and a Method for Prediction”. In: *Technometrics* 16.1, pp. 125–157. DOI: 10.2307/1267500.

Chai, C. P. (2019). “Text Mining in Survey Data”. In: *Survey Practice* 12.1, pp. 1–14. DOI: 10.29115/SP-2018-0035.

Daumé III, H. (2007). “Frustratingly Easy Domain Adaptation”. In: *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pp. 256–263.

Fawcett, T. (2006). “An introduction to ROC analysis”. In: *Pattern Recognition Letters* 27, pp. 861–874.

Friedman, J., R. Tibshirani, and T. Hastie (2010). “Regularization Paths for Generalized Linear Models via Coordinate Descent”. In: *Journal of Statistical Software* 33.1, pp. 1–22. DOI: 10.18637/jss.v033.i01.

Grimmer, J. and B. M. Stewart (2013). “Text as data: The promise and pitfalls of automatic content analysis methods for political texts”. In: *Political analysis* 21.3, pp. 267–297.

Groot de, J. (2018). *Onderzoeksdesign WIN Consent Survey 2018*. Internal CBS report.

Hornik, K., J. Rauch, C. Buchta, and I. Feinerer (2023). *textcat: N-Gram Based Text Categorization*. R package version 1.0-8. URL: https://CRAN.R-project.org/package=textcat.

Keusch, F., B. Struminskaya, C. Antoun, M. P. Couper, and F. Kreuter (2019). “Willingness to participate in passive mobile data collection”. In: *Public opinion quarterly* 83.S1, pp. 210–235.

Pan, S. J. and Q. Yang (2009). “A survey on transfer learning”. In: *IEEE Transactions on knowledge and data engineering* 22.10, pp. 1345–1359.

Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. K. Gadarian, B. Albertson, and D. G. Rand (2014). “Structural topic models for open-ended survey responses”. In: *American journal of political science* 58.4, pp. 1064–1082.

Seni, G. and F. Elder (2010). “Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions”. In: *Synthesis Lectures on Data Mining and Knowledge Discovery* 2.1, pp. 1–126.

Stone, M. (1977). “An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike’s Criterion”. In: *Journal of the Royal Statistical Society, Series B (Methodological)* 39.1, pp. 44–47. DOI: 10.1111/j.2517-6161.1977.tb01603.x.

Straka, M., J. Hajič, and S. J. (2016). “UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing”. In: *LREC 2016 proceedings*. Ed. by N. Calzolari et al. Portorož, Slovenia: European Language Resources Association (ELRA).

Struminskaya, B., P. Lugtig, V. Toepoel, B. Schouten, D. Giesen, and R. Dolmans (2021). “Sharing data collected with smartphone sensors: Willingness Participation and Nonparticipation Bias”. In: *Public Opinion Quarterly* 85.S1, pp. 423–462. DOI: 10.1093/poq/nfab025.

Struminskaya, B., V. Toepoel, P. Lugtig, M. Haan, A. Luiten, and B. Schouten (2020). “Understanding willingness to share smartphone sensor data”. In: *Public Opinion Quarterly* 84.3, pp. 725–759. DOI: 10.1093/poq/nfaa044.

Swets, J. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers. Scientific psychology series*. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Tay, J., B. Narasimhan, and T. Hastie (2023). “Elastic Net Regularization Paths for All Generalized Linear Models”. In: *Journal of Statistical Software* 106.1, pp. 1–33. DOI: 10.18637/jss.v106.i01.

Universaldependencies.org (2014-2024). *Universal Dependencies (UD)*. Framework for consistent annotation of grammar. URL: https://universaldependencies.org/.

Wijffels, J. (2023). *udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit.* R package version 0.8.11. URL: https://CRAN.R-project.org/package=udpipe.

Appendices

A Frequency tables consent survey

Frequency tables of background characteristics of the consent survey used in the analyses.

Table A.1 Frequency of consent-survey response by gender.

	Gender	Frequency	Percent	Valid percent
Valid	male	855	43.5%	43.5%
	female	1110	56.5%	56.5%
	total	1965	100.0%	100.0%

Table A.2 Frequency of consent-survey response by age class.

	Age	Frequency	Percent	Valid percent
Valid	16–29 years	588	29.9%	29.9%
	30–49 years	613	31.2%	31.2%
	50–67 years	530	27.0%	27.0%
	68+ years	234	11.9%	11.9%
	total	1965	100.0%	100.0%

Table A.3 Frequency of consent-survey response by education level.

	Education	Frequency	Percent	Valid percent
Valid	lower	77	3.9%	4.0%
	intermediate	996	50.7%	51.5%
	higher	861	43.8%	44.5%
	total	1934	98.4%	100.0%
Missing		31	1.6%	
total		1965	100.0%	

B Language model

The rule-based language model used for extracting keywords from sentences in the responses consists of the following steps (with example keywords):

1. If *upos parent* = adjective,
   a. select “token and parent” (bigram) if token = “not” (e.g., “not relevant”, “not important”)
   b. else select parent (e.g., “funny”, “personal”)
2. If *upos parent* = noun,
   a. if token = “no” or stem token in {“previous”, “preced”, “preceed”, “same”, “earlier”}, select “token and parent” (bigram) (e.g., “previous answer”, “no problem”)
   b. else select “token and parent” (bigram) if *upos token* = adjective (e.g., “wrong hands”, “high risk”)
   c. else select parent (e.g., “movie”, “questionnaire”)
   d. replace bigrams overlapping between 2a and 2b with a trigram (e.g., “no task” and “difficult task” becomes “no difficult task”).
3. If *upos parent* = verb, distinguish three sentence structures:
   a. “I do not want to”, which results in “not want”,
   b. “I do not find it necessary”, which results in “not necessary”,
   c. “I find it unnecessary”, which results in “unnecessary”.

   Structures 3a and 3b are obtained if token $\in$ {“not”, “never”} and *dep_rel* = adverbial modifier. Structures 3b and 3c are obtained if *upos token* = adjective and *dep_rel* = open clausal complement. The separate selections 3a, 3b, and 3c are obtained by computing the intersection of these sets and the differences between the intersection and the two separate sets. For 3a, “token and lemma parent” (bigram) is selected; for 3b, the bigram “not + token”; and for 3c, the token.
4. Other selections
   a. If the sentence consists of one token, select token (e.g., “privacy”, “instinctively”).
   b. If the sentence consists of two tokens where the second is punctuation, select the first token (e.g., “no”, “n.a”).
   c. If stem token $\in$ {“previous”, “preced”, “preceed”, “same”, “earlier”, “befor”}, select token (e.g., “before”).
   d. If token = “no” and *upos parent* = verb, select “token and parent” (bigram) (e.g., “no value”).
   e. If parent $\in$ {“not”, “never”}, select “token and parent” (bigram) (e.g., “just not”).
   f. If *upos parent* = verb and *dep_rel* = phrasal verb particle, select “token and parent” (bigram) (e.g., “do with”).
   g. If *upos token* $\in$ {adjective, noun}, select token (e.g., “privacy”, “unknown”).
   h. If *upos parent* = verb, select lemma parent (e.g., “want”, “feel”).
   i. If token $\in$ {“not”, “never”} and *upos parent* = auxiliary, select “token and lemma parent” (bigram) (e.g., “not able”).
   j. If *upos token* $\in$ {verb, adverb, pronoun, interjection, other}, select token (e.g., “enough”, “why”).

The selection rules 4a–4j are applied to ensure that for each sentence at least one keyword is selected. These rules were only applied to a limited number of text fields.

C Codebook LISS

Table C.1 Codebook for manually assigned comments in the LISS panel

Category	Description	Examples
Privacy	Privacy-related responses, such as "too personal" or feelings of loss of anonymity.	"I find it unnecessary to share personal data."
Misuse	The respondent does not trust the agency or is afraid that the data will be misused.	"I don't trust handing over data to third parties; this can be misused."
Safety	The respondent is afraid that the data will not be sufficiently protected and will fall into the hands of unauthorised persons.	"These days there is no guarantee that your data is safe."
Effort	It costs the respondent too much time or effort.	"This is going too far for me."
Control	The respondent wants to know what happens to the data, for what purpose it is used, who uses it and how it is used.	"I like to keep the reins in my own hands."
Surveillance	The respondent feels the agency already knows too much about them.	"Everyone already knows far too much about each other."
Elsewhere	The respondent feels the required information is already available elsewhere.	"There are already plenty of other opportunities to obtain the info."
Without reason	The respondent does not want to share information without giving a specific reason and the reason cannot be related to feelings.	"That is nobody's business."
Due to emotions	The respondent does not want to share information because it does not feel right. Reasons relate to emotions or feelings.	"Not a pleasant idea.", "Find it annoying."
Other/don't know	Responses that cannot be classified in the other categories, including "don't know".	"No idea.", "No interest."
Non-informative	Non-informative responses like symbols, characters or punctuation marks.	"-", ".", "...", "n.a."

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Grafimedia

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information