Appendix

A Corpus Collection

A.1 Communicative Discourse

The communicative discourse consists of 6235 newspaper articles collected from 40 newspapers via Factiva and NexisUni (see Table Table 3 for an overview). We have articles from most important quality newspapers across the political spectrum. We could not obtain articles from Frankfurter Allgemeine Zeitung and Libération in particular, but we do have other center-right and center-left newspapers from Germany and France. To be part of our sample, articles had to contain at least one reference to the

gig economy **OR** (automation/robotization/artificial intelligence) **AND** (job **OR** jobs **OR** work) **OR** fourth industrial revolution **OR** industry 4.0.

The specific search strings for the non-english-speaking countries were as follows:

France: [gig economy **OR** uberisation] **OR** [[travail or emploi] **AND** [robotisation **OR** automatisation **OR** numérisation **OR** Intelligence Artificielle]] **OR** [quatrième* révolutio* industriell* **OR** Industrie 4.0]

Germany: gig economy **OR** [[arbeit or arbeitspl*] **AND** [automatisierung **OR** robot* **OR** künstliche* intelligenz]] **OR** [viert* industr* revolution **OR** Industrie 4.0]

Italy: gig economy **OR** [lavoro **AND** [intelligenza artificiale **OR** automazione **OR** robotizzazione]] **OR** [Industria 4.0 **OR** Impresa 4.0 **OR** Quarta rivoluzione industriale]

Poland: gig economy **OR** [[prac* **OR** zatrudnieni*] **AND** [automatyzacj* **OR** robotyzacj* **OR** sztuczn* inteligencj*]] **OR** [przemysł 4.0 **OR** platform* przemysłu przyszłości **OR** Czwart* rewolucj* technologic*]

Spain: gig economy **OR** [trabajo **AND** [automatización **OR** robotización **OR** inteligencia artificial]] **OR** [Industria 4.0 **OR** cuarta revolución industrial]

Sweden: [gig-ekonomi **OR** gig-jobb **OR** gig economy] **OR** [[jobb or arbete] **AND** [automation **OR** automatisering **OR** robot* **OR** artificiell intelligens]] **OR** [Industry 4.0 **OR** industri 4.0 **OR** fjärde industriella revolutionen]

To make it more likely that articles were really about these topics and did not just contain a single mention somewhere in an otherwise unrelated text, these search strings had to appear in the headline or the lead paragraph of articles. The choice of search strings was meant to ensure that the discourse on the digital future of work is represented in its entirety. In countries where gig economy is uncommon and did not yield many results, we included equivalent language-specific terms such as ubérisation in France. For each country, we tried different combinations of search strings and manually checked whether they work, i.e.: do they result in a number of articles comparable, given the country's size, to other countries? Do they return articles that cover what we are interested in?

Another potential problem is that the term industry 4.0 is relatively specific (but not exclusive) to some countries (Germany, Italy). We therefore also included the term fourth industrial revolution (in addition to robotization and automation), which should ensure that a higher proportion of industry reference really is due to the fact that these countries talk more about industrial issues – and not due to the fact that there is a terminological bias in our sample. A final problem is that the term gig economy is used more broadly in the UK (and Ireland), and not just referring to digital labor platforms. We do want to capture when gig work and digital gig work are discussed together, but we do not want articles that just talk about gig work in ways that articles in other countries would talk about precarious work. To remedy this, we removed articles that did not contain at least one reference to specifically digital aspects of the gig economy (such as Uber, digital platform, etc.). We also skimmed over all articles and manually removed those that were still not really related to our topic of interest. In a last step, we removed very long (>30000 characters) and very short (<200 characters) articles. All in all, our goals was to make the articles included in our sample as relevant as possible. This means much less noise in our data than had we simply tried to maximize the number of articles based on relatively general search strings.

Table 3: Newspaper Corpus Overview

Newspaper	Newspaper Leaning	Number of Articles
France		
L'Humanité	(center-)left	93
Les Echos	(center-)right	389
Le Figaro	(center-)right	119
La Croix	(center-)right	30
Le Monde	center-left	138
L'Opinion	centrist	85
Capital Finance	centrist	12
Germany		
Süddeutsche Zeitung	(center-)left	233
Der Tagesspiegel	(center-)left	135
Deutschlandfunk	(center-)left	76
DIE ZEIT	(center-)left	45
taz - die tageszeitung	(center-)left	33
Die Welt	(center-)right	239
Handelsblatt	centrist	466
Ireland		
The Irish Times	(center-)left	103
Irish Independent	(center-)right	42
The Irish Examiner	centrist	24
Italy		
La Repubblica	(center-)left	264
Il Sole 24 Ore	(center-)right	400

Table 3: Newspaper Corpus Overview *(continued)*

Newspaper	Newspaper Leaning	Number of Articles
Corriere della Sera	centrist	656
La Stampa	centrist	103
Poland		
Rzeczpospolita	(center-)right	290
Gazeta Wyborcza & Wyborcza.pl	centrist	109
Gazeta.pl	centrist	44
Dziennik Gazeta Prawna	centrist	24
Wyborcza.biz	centrist	18
Spain		
El País	(center-)left	330
El Periódico	(center-)left	93
La Vanguardia	(center-)right	194
ABC	(center-)right	93
El Mundo	centrist	81
Sweden		
Aftonbladet	(center-)left	23
Svenska Dagbladet	(center-)right	100
Dagens Nyheter	centrist	190
Business Post	centrist	80
United Kingdom		
Guardian/Observer	(center-)left	264
Telegraph/Sunday Telegraph	(center-)right	245
Financial Times	centrist	191
Independent/Independent on Sunday	centrist	181

A.2 Coordinative Discourse

Our coordinative discourse consists of 2337 documents (see Table 4 for an overview). To enter our corpus, documents had to directly deal with either digitalization of production processes (automation, robotization, Industry 4.0) or new, digitally-enabled forms of work (platform economy, gig work). While the bulk of documents are standalone, some of them are the result of splitting apart longer documents with several sub-chapters. The documents were published between 2012 and 2019 (see Figure 4 for an overview).

Documents were manually collected from the websites of policy actors. We focused on the major social partners, namely governments, employer organizations, and trade unions. The government category includes relevant ministries such as those charged with innovation, work, technology, or social policy. Generally, we selected employer organizations that are part of Business Europe. But since our goal is to obtain documents from the most representative policy actors in a country, we also used – when available – qualitative data from DIRESOC Project, whose country reports provide important information about the importance of specific social partners. When an organization did not belong to Business Europe, but was mentioned by DIRESOC country reports, we included it. We also gave centrality to representativeness when selecting trade unions.

We chose unions that are part of the European Trade Union Confederation (ETUC). Following the same logic as in the case of employer organizations, we triangulated this criterion with DIRESOC country reports. In addition, if actors were considered relevant based on country-specific knowledge but were not part of our sample so far, we still included them as in the case of the German IGMetall or the British IWGB. Due to linguistic barriers, we asked to native speakers to collect policy documents on Poland and Sweden. We asked them to follow the guidelines described above.

Table 4: Policy Documents Corpus Overview (Note that documents are of unequal length).

Actor	Actor Type	Number of Documents
France		
CFDT	Union	7

Table 4: Policy Documents Corpus Overview (Note that documents are of unequal length). *(continued)*

Actor	Actor Type	Number of Documents
CGT	Union	20
CNN	Government	123
COE	Government	22
CPME	Employer	6
FO	Union	69
Government	Government	15
MEDEF	Employer	4
MEF-MRP	Government	27
MEFI-MBCR	Government	54
MT	Government	8
Socialpartners	Union	16
U2P	Employer	11
Germany		
BDA	Employer	12
BDI	Employer	5
BMAS	Government	66
BMBF	Government	11
BMWi	Government	24
BR	Government	3
BVMW	Employer	1
DGB	Union	62
IGBCE	Union	28
IGMetall	Union	19
INQA	Government	1

Table 4: Policy Documents Corpus Overview (Note that documents are of unequal length). *(continued)*

Actor	Actor Type	Number of Documents
N3etwerk	Employer	1
VDMA	Employer	1
verdi	Union	18
Ireland		
DBEI	Government	34
DCENR	Government	9
IBEC	Employer	19
ICTU	Union	2
Italy		
CGIL	Union	42
CGILCISLUIL	Union	1
CISL	Union	8
Confindustria	Employer	157
FDV	Union	4
INAPP	Government	9
MISE	Government	7
MLPS	Government	2
Poland		
BCC	Employer	5
GovernmentFPPP	Government	1
GovernmentMC	Government	8
GovernmentMinisters	Government	2
GovernmentMPT	Government	7
GovernmentMRPPS	Government	1

Table 4: Policy Documents Corpus Overview (Note that documents are of unequal length). *(continued)*

Actor	Actor Type	Number of Documents
GovernmentV4	Government	1
Lewiatan	Employer	13
NSZZSolidarnosc	Union	4
OPZZ	Union	4
PracodawcyRP	Employer	6
Spain		
CC	Employer	6
CCOO	Union	50
CEOE	Employer	76
Cepyme	Employer	348
ConsEcoSoc	Government	12
F1M	Union	39
MEE	Government	13
MEETA	Government	52
MESS	Government	46
MIET	Government	46
MIET-MHAP	Government	7
MITC	Government	12
MTMSS	Government	23
UGT	Union	45
Sweden		
Gov	Government	297
LO	Union	15
SACO	Union	56

Table 4: Policy Documents Corpus Overview (Note that documents are of unequal length). *(continued)*

Actor	Actor Type	Number of Documents
SvensktNar	Employer	9
TCO	Union	38
United Kingdom		
CBI	Employer	11
Commons	Government	6
DDCMS	Government	11
FSB	Employer	24
Gov	Government	32
IWGB	Union	4
Taylor	Government	14
TUC	Union	65

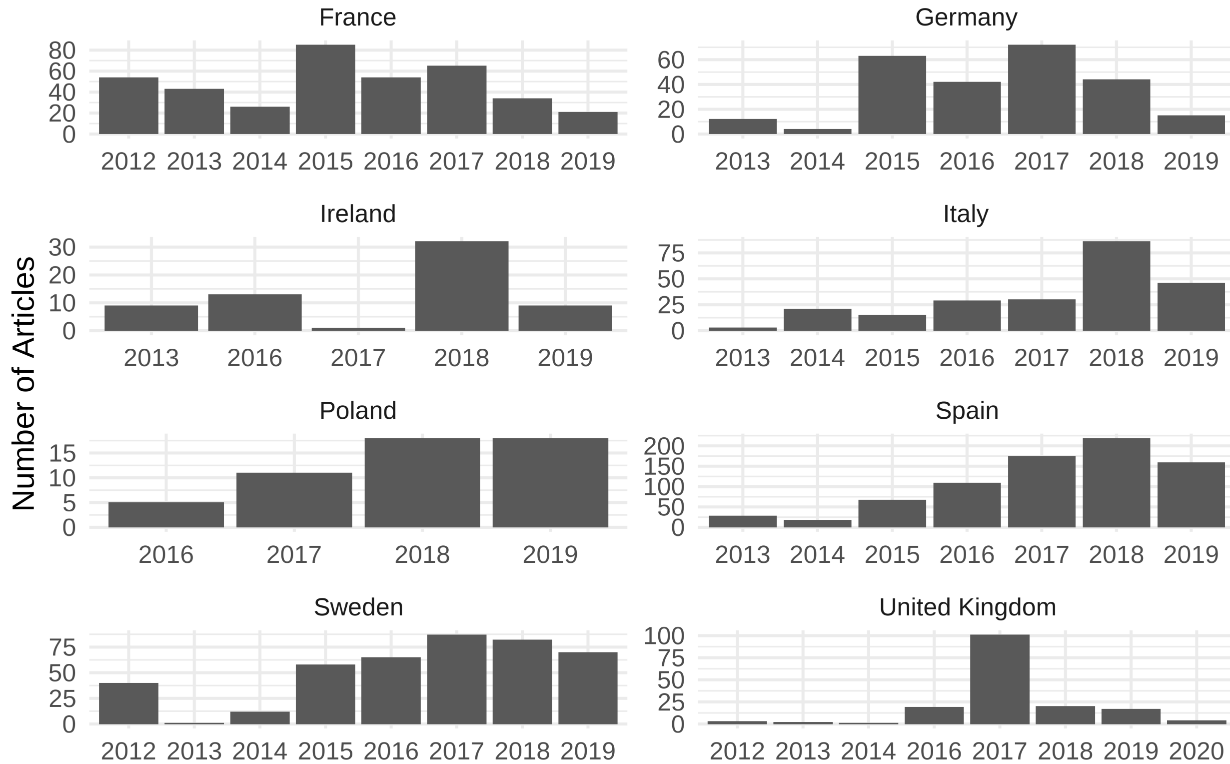

Figure 4: Number of Documents per Year (Note that documents are of unequal length).

B. Details on Methods

B.1 Sentiment Analysis

In the paper, we use the Lexicoder Sentiment Dictionary developed by Young and Soroka (2012). While the dictionary was specifically designed for political text and is therefore be appropriate for our corpora, we cross-validated our findings with other dictionaries given the validity issues associated with dictionary-based analysis (Boukes et al. 2020; van Atteveldt, van der Velden, and Boukes 2021). For these dictionaries, we complemented the dictionary-based approach to sentiment analysis with natural language process in order to allow us to account for negators, amplifiers and deamplifiers. We used standard negators, amplifiers and deamplifiers obtained from the lexicon package (Rinker 2018). Specifically, we used the following four widely used and relatively general dictionaries to cross-validate our findings:

– the AFINN dictionary developed by (Nielsen 2011);

– the Bing sentiment lexicon developed by Hu and Liu (2004);

– the NRC Word-Emotion Association Lexicon developed by Mohammad and Turney (2010); and

– the syuzhet dictionary (and accompanying R package) developed by Matthew Jockers and the Nebraska Literary Lab (Jockers 2017).

In case of the NRC dictionary, we did not use their emotion-based dictionaries (such as those for anger, fear, or hope) but only those for positive and negative terms. The afinn and syuzhet dictionaries have a higher resolution than bing and nrc by scoring words not just as negative (-1) and positive (+1) but allowing more gradation. We transformed these higher-resolution scales to a binary positive-negative score for both technical reasons and to provide more robust estimates. After all, whether a word is positive or negative is a much more straightforward question than whether a positive word is quite (+3), very (+4) or extremely (+5) positive, especially across different contexts and for machine-translated documents.

The strong similarities across dictionaries make us confident that the differences we measure are real differences and not driven by the particularities of any one of these dictionaries (see Figure 4). This is particularly true for the larger differences, i.e., the bottom and top of the distribution, on which we focus in the paper (we don't, for example, substantively discuss the small differences between Germany, France and Spain).

Figure 5 shows the 50 most frequently matched positive and negative words for the different dictionaries. This is meant to both illustrate what positive and negative tone mean and to validate the dictionaries. As is evident from the plot, the most commonly matched words are what one would expect to appear in positive and negative framings of the digital future of work.

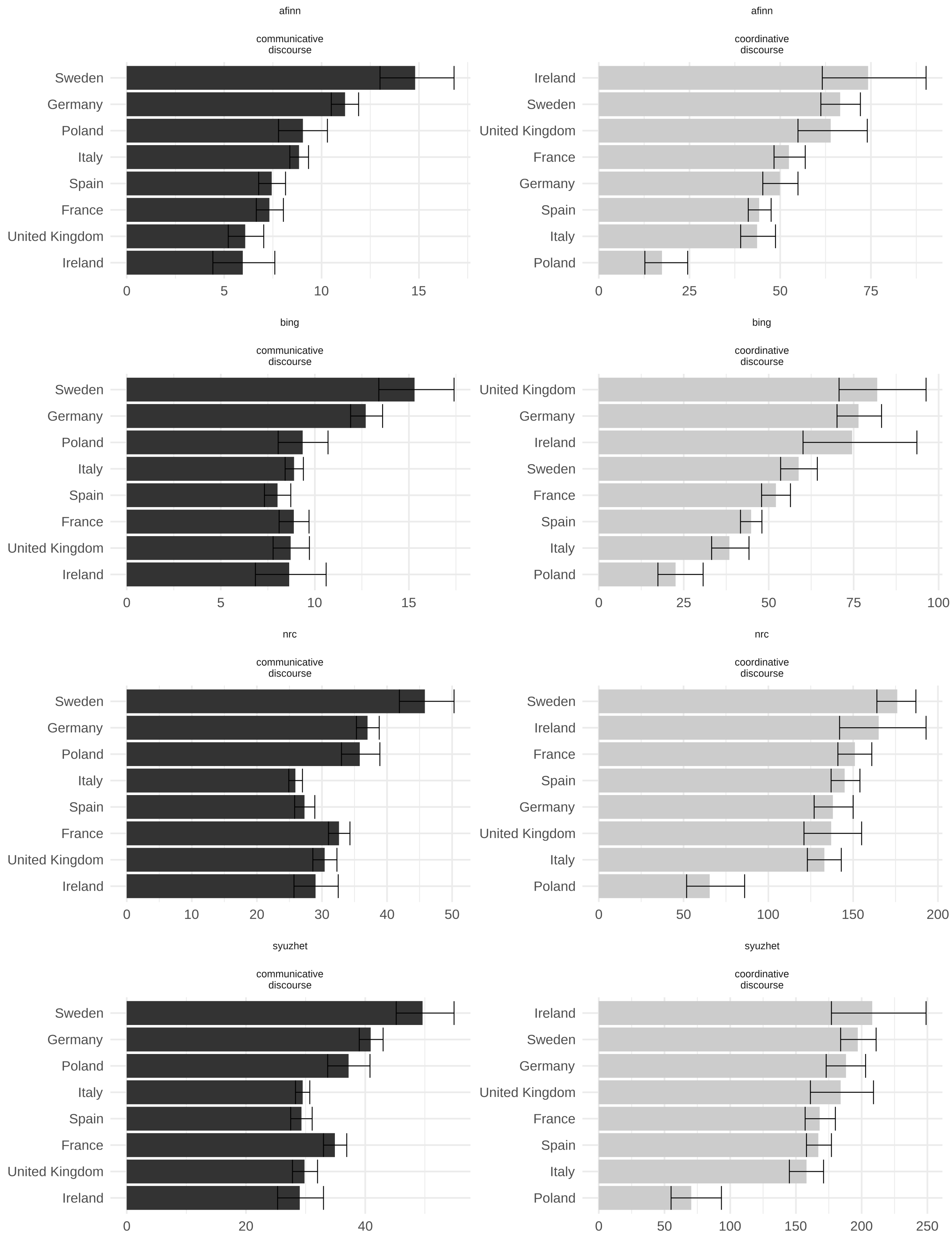

Figure 5: Discursive sentiment across countries, dictionaries and discourse types with 95 % bootstrapped confidence intervals (R = 10000)

Figure 6: Wordclouds with the most commonly matched negative and positive words

B.2 Keyword Extraction Techniques

RAKE The RAKE – or rapid automatic keyword extraction – algorithm starts from the idea that keywords usually contain several informative words bur rarely punctuation or stopwords (Rose et al. 2010). It thus first tokenizes a given text, using spaces and punctuation to break it at word delimiters. It then creates sequences of contagious words, so called candidate keywords. For example, the short text "robots and artificial intelligence will change the nature of work and our work life balance" will be tokenized as follows:

["robots," "and," "artificial," "intelligence," "will," "change," "the," "nature," "of," "work," "and," "our," "work," "life," "balance"]

Reading from left to right, the algorithm then creates candidate keywords every time a common stopword is encountered, like this

["robots," "artificial intelligence," "change," "nature," "work," "work life balance"]

Next, a score is created using the following formula:

$$wordscore = \frac{degree(word)}{frequency(word)}$$

Frequency refers to the number of times a word appears in the list of candidate keywords.

$frequency(robots) = 1,$

$frequency(artificial) = 1,$

$frequency(intelligence) = 1,$

$frequency(change) = 1,$

$frequency(nature) = 1,$

$frequency(work) = 2,$

$frequency(life) = 1,$

$frequency(balance) = 1$

Degree refers to how frequently a word co-occurs with other candidate keywords in a given text. This is equivalent to the number of times a word occurs in the candidate keywords that contain this word (a higher word degree can therefore also indicate that it appears in longer candidate keywords):

$frequency(robots) = 1,$

$frequency(artificial) = 2,$

$frequency(intelligence) = 2,$

$frequency(change) = 1,$

$frequency(nature) = 1,$

$frequency(work) = 4,$

$frequency(life) = 3,$

$frequency(balance) = 3$

As we saw, the word score is proportional to the degree of word and inversely proportional to its frequency. RAKE thus favor words that occur not too frequently but often in combination with other keywords, especially long ones. To calculate the candidate keyword score, we add the word scores of its constituent parts and take a the highest-scoring percent (e.g. the highest-scoring 33%) as our keywords to be extracted. For our example, this would look as follows:

$score(robots) = 1/1 = 1,$

$score(artificialintelligence) = (word_score(artificial) = 2/1 = 2 + word_score(intelligence) = 2/1 = 2) = 4$

$score(change) = 1/1 = 1,$

$score(nature) = 1/1 = 1,$

$score(work) = 4/2 = 2,$

$score(worklifebalance) = (word_score(work) = 4/2 = 2 + word_score(life) = 3/1 = 3 + word_score(balance) = 3/1 = 3) = 8$

This means that “work life balance” and “artificial intelligence” were selected as our most central keywords. Rake was implemented using the UDPipe R package (Straka and Straková, 2017).

Co-occurrences We also constructed bigrams based on their number of co-occurrences, either within a skipgram of size four or within a sentence. In the former case (reported in the paper), we counted bigrams when words followed one another directly or when we skipped up to three words in between. In the latter case (reported here), we counted bigrams when they occurred within the same sentence. The results are largely similar to those based on skipgrams, although produce somewhat more sensible results since the sentence restrictions is less sensible as the skipgram-context restrictions (as the latter is narrower).

Textrank Due to space constraints, we did not include Textrank-based keywords in the paper. However, as Figure shows, the words with the highest PageRank score support our general findings. Textrank is a graph-based ranking model that identifies keywords by constructing a word network based on whether two words follow one another (or co-occur in a window of N words) (Mihalcea and Tarau 2004). If they do, an edge is created between them, the weight of which depends on how often they follow each other in a given text. Next, the PageRank algorithm is applied to this network to rank words in their order of importance (this is what is reported in Figure). Relevant words that follow one another can then be combined to obtain keywords. Textrank was implemented using the R textrank (Wijffels 2019).

Figure 7: Keywords with the highest PageRank score

B.3 Word Vectors

In addition to the analysis presented in the paper, we also use word vectors to identify semantic similarities among words and to assess whether they systematically differ across countries for key terms of interest. Word vectors are based on the idea that words that appear near each other have similar meanings. Or, in John Rupert Firth famous phrase, we 'know a word by the company it keeps' (Spirling and Rodriguez 2019). Words are mapped to a multidimensional vector that represents its 'meaning,' with the different dimensions representing different semantic or syntactic connotations of a word. Somewhat unorthodoxly, we used simple word counts and matrix factorization to calculate word vectors (Moody 2017).

We used a moving slide window to create skipgrams of length eight and then calculated the pointwise mutual information (PMI) metric for all word pairs (see, Silge 2017). This gives us information about which words occur together (within a moving window of size 8) more often than expected based on how often they occur on their own. PMI is simply the logarithm of the normalized skipgram probability which in turn is the result of dividing the frequency of two words co-occurring in a skipgram by the frequencies of their individual occurrence. This is a well-understood metric for how frequently two words occur jointly rather than independently (Moody 2017). We then cast a sparse matrix where each row represents word 1 and each column word 2 and each value is the PMI calculated above. From there, we reduce the dimensionality of that matrix using singular value decomposition, specifying the number of principal components to 300 – a value common in the word embeddings literature. Using a tidy approach, we now have a tibble of word vectors which we can use to find synonyms (or analogues, etc.).

Here, we first provide additional examples for this synonym task which demonstrate that our approach reliably finds highly plausible synonyms. As Table 5 shows, our approach works surprisingly well given that it is much less complex and computationally demanding than using neural networks, and given the relatively small size of the data. We can then do the same for terms that are of more theoretical interest. Table 6 depicts a number of such words. Basically, what we do here is to zoom in on words of interest by looking at nearby vectors in a vector space, that is, by looking at synonyms. We see, for example, that in Italy and Germany terms like '4.0' or factory are closely associated with future and work, reflecting the importance for both countries of getting digital manufacturing right. In line with what we found in the paper, in German discourse the future is viewed in more positive terms, with work-life-balance issues taking up a prominent place (time, family, unpaid). The French discourse, to give a second illustrative example, reflects the political commitment to making France a digital frontrunner, and the concomitant entrepreneurial role ascribed to the state in general and the National Digital Council (Conseil National du Numérique) in particular. Accordingly, words like project, program and plan are associated with the future, and words like ambition, national, or public are associated with digital. A final example are the UK and Ireland where work is closely associated with insecurity, reflecting the often-precarious nature of these countries' labor markets.

Table 5: Synonyms for capital, gates, google, and public

country	capital	gates	google	public
France	capital, venture, funds, fund, investment, ventures, partners, private, equity, angels, risk, seed, fcpr, stage, scr	gates, musk, elon, bill, stephen, hawking, boss, ne, pas, tesla, bezos, ou, late, etats, spacex	google, apple, facebook, amazon, search, microsoft, 2012, maps, engine, twitter, gmail, hotel, youtube, beat, 2010	public, innovation, procurement, private, services, data, service, action, digital, economy, support, policies, social, purchasing, participation
Germany	capital, venture, investment, fund, wealth, investors, start, ups, money, profits, short, investments, taxes, income, invested	gates, founder, boss, bill, tesla, microsoft, day, musk, elon, fair, cyberattacks, thousands, shareholders, hanover, march	google, facebook, apple, amazon, microsoft, giants, valley, alphabet, search, corporations, alexa, silicon, uber, pichai, ibm	public, administration, service, transport, services, scientific, insurance, private, defense, wlan, million, social, financial, sector, debate
Ireland	capital, allowances, venture, relief, accelerated, tax, credit, assets, investors, funding, efficient, start, expenditure, ventures, investment	gates, bill, founder, musk, elon, microsoft, apple, david, william, attendees, german, steve, stephen, japanese, phrase	google, facebook, amazon, translate, search, twitter, linkedin, ebay, youtube, recognition, marketplaces, siri, card, offline, marshall	public, sector, services, consultation, private, healthcare, service, administration, bodies, community, security, office, 1.1, costs, national
Italy	capital, human, investments, venture, equity, investment, foreign, funds, tangible, fixed, assets, intangible, loans, banks, depreciation	gates, bill, elon, musk, founder, stephen, hawking, robots, tesla, robot, techno, physicist, fear, income, including	google, facebook, amazon, microsoft, giants, intelligence, artificial, apple, bezos, glass, silicon, valley, jeff, ceo, cars	public, private, administration, investment, financial, national, research, funds, intervention, finance, policies, support, debt, social, investments
Poland	capital, funds, german, foreign, venture, investors, deloitte, investment, fund, human, vc, startup, financed, ventures, investor	gates, bill, stephen, elon, musk, fears, hawking, previously, shared, race, expressed, warned, late, contact, famous	google, duplex, assistant, pichai, facebook, sundar, giant, tesla, amazon, apple, ai, conversation, alphabet, wave, presentation	public, administration, sector, private, entities, entity, information, consultations, disclosure, procurement, official, inter, isp, governmental, website

Table 5: Synonyms for capital, gates, google, and public *(continued)*

country	capital	gates	google	public
Spain	capital, venture, barcelona, ventures, madrid, human, club, corporate, 250k, 1m, seed, 500k, 2m, fund, 5m	gates, founder, bill, microsoft, rowe, philanthropist, ii, iii, farming, richard, robots, price, page, firm, vertical	google, amazon, microsoft, apple, facebook, samsung, huawei, artificial, alphabet, tecnologia, twitter, hp, intel, intelligence, netflix	public, private, administration, administrations, bodies, services, system, participation, institutions, collaboration, information, investment, agencies, entities, policies
Sweden	capital, venture, human, financing, almi, inland, loan, invest, guarantees, gains, subtotal, sme, productivity, private, investments	gates, founder, bill, musk, elon, steven, recently, tesla, silicon, valley, hawking, ingvar, stephen, physicist, henry	google, microsoft, facebook, amazon, ibm, ericsson, google's, apple, behshad, behzadi, news, giants, search, volvo, investing	public, sector, private, procurement, business, service, authorities, publications, actors, activities, transport, administration, customers, publicly, services
United Kingdom	capital, venture, labour, funds, investment, equity, tech, investors, fund, markets, firm, deals, wealth, closed, trading	gates, bill, elon, musk, stephen, hawking, founder, mark, zuckerberg, warned, billionaires, www.independent.co.uk, tax, ceo, pizza	google, facebook, google's, deepmind, microsoft, search, google's, siri, amazon, apple, alphabet, schmidt, deep, owned, bought	public, sector, local, services, private, policy, libraries, radio, content, investment, finances, england, health, report, test

Table 6: Word vector similarities for different keywords across countries

country	digital	future	work
France	digital, economy, ambition, innovation, technology, companies, public, council, society, literacy, france, national, french, appendices, sectors	future, industry, project, factory, tomorrow, program, education, european, world, labor, industrial, investments, impact, launched, plan	digital, time, economy, life, employees, workplace, employment, skills, information, mission, home, workers, labor, autonomy, technologies
Germany	digital, industry, digitization, 4.0, change, world, data, transformation, employees, capitalism, platforms, technology, future, virtual, business	future, 4.0, industry, factory, digital, world, people, digitization, change, development, machines, social, german, market, germany	digital, world, employees, family, future, 4.0, time, life, industry, intensity, people, digitization, unpaid, hours, forms

Table 6: Word vector similarities for different keywords across countries *(continued)*

country	digital	future	work
Ireland	digital, transformation, economy, internet, ai, market, technologies, single, related, business, workers, technology, innovation, infrastructure, ecosystem	future, expert, study, legal, market, systems, technologies, ability, world, 1, current, ibec, europe, trends, 27	insecure, digital, life, contracts, hours, smarter, development, one's, plan, scope, appendix, flexible, forms, insight, placements
Italy	digital, innovation, 4.0, industry, hubs, transformation, hub, technological, technologies, economy, dih, platforms, industrial, business, manufacturing	future, automation, factory, robots, revolution, interventions, people, 4.0, frey, osborne, 1, oxford, industry, press, articoli	digital, people, organization, workers, employment, production, hours, 4.0, training, labor, paper, forms, tasks, agile, world
Poland	digital, transformation, competences, innovation, economy, competence, technologies, solutions, priority, kazimierz, talents, industrial, identified, business, i.e	future, foundation, report, professions, platform, job, industry, labor, en, act, automation, field, jobs, 4.0, offers	intelligence, employees, people, artificial, automation, services, repetitive, data, routine, physical, employee, performing, actions, poland, labor
Spain	digital, industry, transformation, spain, digitization, 4.0, business, economy, companies, technologies, sector, ametic, plan, society, technology	future, author, lab, accelerator, revolution, market, bic, seedrocket, current, challenges, ventures, industrial, robots, initland, life	workers, life, hours, health, organization, jobs, people, environment, home, time, hygiene, labor, psychosocial, assessment, business
Sweden	digital, agenda, platforms, competence, services, digitization, skills, service, market, regional, economy, commission, 2015, agendas, development	future, society, commission, digitalisation, digitization, challenges, report, intelligence, space, employment, skills, artificial, world, researcher, competence	environment, digital, labor, performed, tasks, economy, market, time, tax, employment, responsibility, 24, business, platforms, workplace
United Kingdom	digital, skills, businesses, uk, technology, business, connectivity, firms, dcms, smes, infrastructure, world, future, voice, plans	future, digital, automation, http, skills, jobs, shaping, report, 5g, world, economy, infrastructure, digitalisation, tech, intelligence	people, jobs, insecure, hours, world, economy, time, life, job, balance, future, technology, employment, activities, gig

B.4 Topic Modelling

Preprocessing For pre-processing, we used annotated part-of-speech tags to select nouns, adjectives, and verbs. We discarded punctuation and stopwords as well as semantically less meaningful parts of speech like determiners or names entities like dates. This seems justifiable, given that we are not interested in linguistic style or subtle word use but rather in the broad thematic contours of discourse. We also discarded location-specific information like capital cities or languages to avoid linguistically (as opposed to substantively induced) country-effects.

We also constructed a list of frequent n-grams such as artificial intelligence, machine learning, virtual reality, big data, tax evasion, further training etc. This list, which contains 82 n-grams, was manually compiled based on the most frequent collocations identified in the text (with log-frequency biased mutual dependency used as the ordering metric). We see no reason not to include such information as there are obvious theoretical reasons to prefer such n-grams to their separate constitutive unigrams.

We lowercased but did not stem our document feature matrix as the difference between singular and plural forms can be meaningful while the difference between uppercased and lowercased words is most likely not – at least in the types of policy and newspaper documents we are looking at. For example, it can make a difference whether a text speaks of robots in the plural – as in the abstract threat that robots pose ('the robots will take our jobs') – and a robot in the singular, which is more likely to be described as something useful or positive ('the robot does x').

We also removed remaining word trash such as html tags, common untranslated words (e.g. della), as well as country-specific information using the named entity information. This latter removal is meant to ensure that differences in topic prevalence are, as much as possible, the result of substantive differences and not of local vernaculars or parochial word usage.

Lastly, we removed words that appeared in more than 50% of documents as such words do not contain much information (e.g. digital). We also removed words that appeared in less than 0.5% of documents. While this is a somewhat arbitrary (although commonly used) standard, qualitative inspection revealed but proofed to be a useful threshold that removed many very specific and rare terms while still retaining un-common but not unimportant words.

Number of Topics We chose a topic model with k = 60 topics. While our decision was assisted by several metrics, the choice was ultimately a theoretical one, based on two criteria. First, given that we are primarily interested in comparing the content of discourse, we want our topics to be broad enough to be at least potentially relevant in different countries, but narrow enough to capture the themes we are interested in (e.g., automation vs compensation). On the one hand, we want obtain topics that are broad enough to be at least potentially relevant in different countries and for different actors. If we chose k=350, for example, we might get many topics that are about particular events in a country (e.g. the introduction of a new technology at a particular company), and will most likely not be discussed in other countries. On the other hand, we want enough topics to allow for meaningful differences to emerge. If we only had, say, 5 topics, these will be too coarse to say anything interesting about cross-country or other differences. Second, while topics are often considered "as an operationalization of policy frames" (Gilardi, Shipan, and Wüest 2020, 3), topic models easily uncover more sensible topics than there could possibly be frames (Nicholls and Culpepper 2020, 8). We therefore opted for using a higher number of topics than we would have had we assumed that topics directly capture real-world frames. We then aggregated these topics into "frame packages" (Nicholls and Culpepper 2020, 11) that capture discursive foci that are neither too broad nor too narrow and are therefore theoretically interesting. For example, topics on industry 4.0, the technological transformation of the production process, cloud computing, or smart factories were combined into the frame package 'digital manufacturing,' which covers debates on how digitalization is changing industrial production and manufacturing.

Based on our intuition about which k should yield topics that are both broad and interesting, we wanted k to be somewhere between 25 and 75. We therefore run topic models with ks between 10 and 100 so as to allow various quantitative metrics to guide our decision. Figure 8 plots four metrics – semantic coherence, exclusivity, residuals, and held-out-likelihood – for models with different ks. Semantic coherence is a metric that measures how often the most frequent words in a topic actually co-occur in a document. While semantic coherence has been shown to correlate well with human judgments of topic quality, it has been shown to increase when topics are dominated by very common words (Roberts, Stewart, and Tingley 2018). Exclusivity, by contrast, penalizes models with few dominant top words. It measures the share of top words which are distinct to a given topic, thus creating something of a trade-off with semantic coherence.

The residuals capture overdispersion of the variance of the multinomial in stm's data generating process (Roberts, Stewart, and Tingley 2018). Higher values indicate overdispersed residuals, implying that the latent topics cannot account for the overdispersion and more topics may be needed to use up the extra variance. Held-out likelihood estimates the probability that words appear in a document when these words have been removed before the estimation. It is a measure of predictive performance, with higher values indicating better performance.

Hence, we want semantic coherence, exclusivity, and held-out likelihood to be as high and the residuals to be as low as possible. Topics numbers higher than the number of plausible individual frames on digitalization - of which there can only be so many - are justifiable as we can combine different topics to frame packages (Nicholls and Culpepper 2020). In the end, we settled for 60 topics, which the quantitaive metrics supported and which also made sense qualitatively. It has to be said, though, that topics were very similar and stable with 10 or even 15 fewer or more topics.

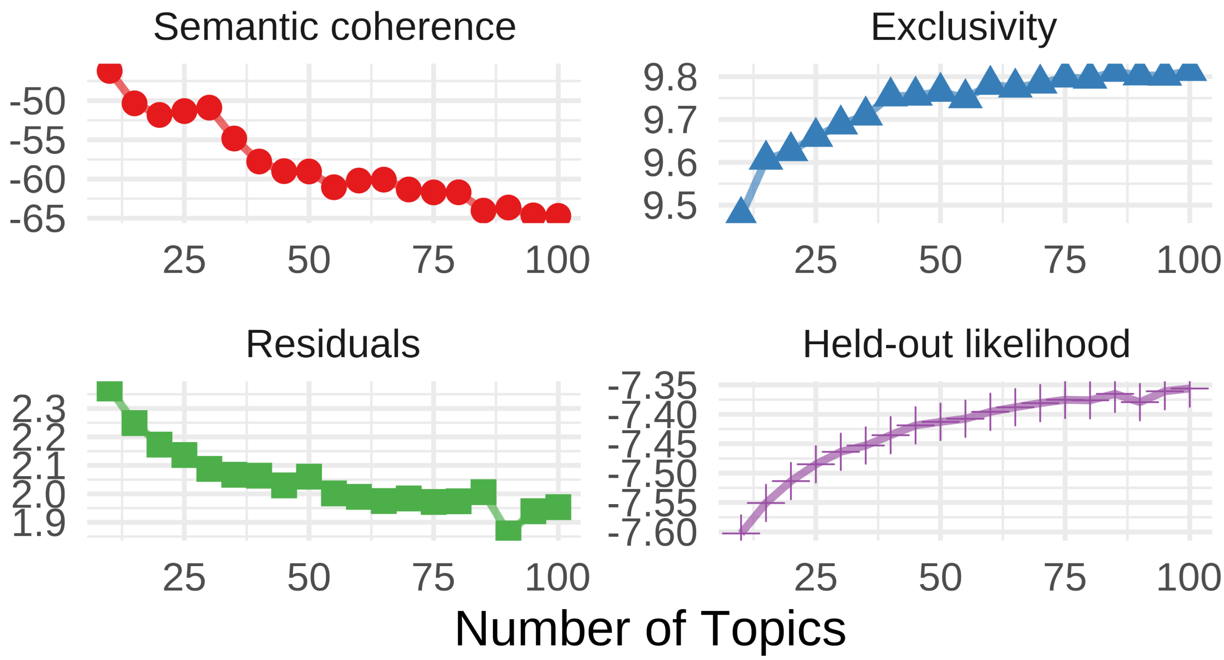

Figure 8: Model diagnostics for different numbers of topics (10 to 100)

Validation

While topic models are powerful tools "for discovering and exploiting the hidden thematic structure in large archives of text" (Blei 2012, 82) they are no magic bullet and need to be carefully interpreted and validated. Validation requirements are somewhat lower if the main goal is not to measure some pre-specified latent trait but, as is the case here, to explore and usefully summarize texts in order to facilitate comparisons (cf. Ying, Montgomery, and Stewart 2019, 1–2). It is nonetheless crucial to at least make this process transparent and to ideally also establish "semantic validity" (Quinn et al. 2010, 216; Ying, Montgomery, and Stewart 2019, 6) by showing that topics are coherent and make sense to external annotators.

Table 7 shows a complete list of topics and, when applicable, the frame package to which we assigned this topic. We only assigned topics to one of the 6 frame packages presented in the paper, which does not mean that the others could not have been assigned to different frame packages or that other topic-frame-package combinations would not have been possible. The labeling of topics as well as their assignment to frame packages was an interpretative process, based on the most *FR*equent and *EX*clusive Words (FREX) and theoretical reasoning. For this reason, we make the topic terms transparent, and we externally verified the semantic validity of our topics. That is, we checked whether they make sense to human readers and whether they can be semantically distinguished from other topics by asking two external researchers to complete two tasks (Chang et al. 2009; Ying, Montgomery, and Stewart 2019). First, we gave them the list below but instead of only the correct category label, we gave the correct label plus two other randomly chosen labels. The task was to pick the correct label. Second, we gave them a list with the top 4 FREX terms plus the highest FREX term from another randomly selected topic. Their task was to correctly identify the 'incorrect' word. In the first task, they, on average, picked the right topic label 85 % of the time; in the second task, they chose the non-fitting word 73,3 % of the time. These results compare well with results from similar tasks (Ying, Montgomery, and Stewart 2019) and make us confident that our topic model produces semantically valid topics.

Table 7: Complete List of Topics and Topic Categories

Topic	Top FREX Words	Topic Label	Aggregated Topic Category
1	artists, admission, story, father, artist, love, art, scene, journalist, film, writer, woman, stories, watch, novel, book, friend, exhibition, journalists, famous	Art	/
2	ambition, inclusion, opening, actors, mediation, open, public, literacy, communities, citizens, practices, citizen, empowerment, administrations, encourage, governance, uses, stakeholders, recommends, administrative	Digital Government	Digital Industrial Policy
3	exposure, prevention, occupational, accidents, preventive, safety, psychosocial, risks, accident, chemical, healthy, exposed, dangerous, nº, substances, protective, agents, diseases, fax, aspects	Occupational Safety	Status Platform Workers
4	industrial, industry, revolution, manufacturing, automotive, production, factories, transformation, smart, industries, advanced, leading, energy, robotics, chains, electronics, competitiveness, strategy, 3d_printing, manufacturers	Industry 4.0	Digital Manufacturing
5	revenue, waste, circular_economy, recycling, materials, environmental, packaging, raw_materials, recovery, circular, reuse, emissions, material, water, disposal, collection, energy, carbon, sustainability, renewable	Sustainability	/
6	patients, patient, healthcare, medical, care, hospital, health, medicine, doctors, doctor, hospitals, benefiting, clinical, drug, treatment, diseases, disease, treatments, drugs, cancer	Health Care	/
7	artificial_intelligence, ai, machine_learning, algorithms, intelligence, algorithm, artificial, computer, language, scientists, ethical, computers, ethics, deep, humans, machine, recognition, neural, images, human	Artificial Intelligence	Automation & Compensation
8	differences, table, digitization, index, percentage, degree, proportion, significant, size, family, smes, workload, intensity, difference, higher, scope, affected, expectations, compatibility, variables	Statistics	/
9	trades, uberization, salaried, collaborative_economy, social_dialogue, collective, subordination, mobilization, impacts, cooperative, status, telework, branch, anticipate, profession, generalization, autonomy, disconnect, subject, evolve	Platform Work	Status Platform Workers
10	taxi, ordering, accounting, license, permit, traffic, center, centers, amendment, vehicle, section, authority, holders, shall, vehicles, equipment, regulations, reservations, licenses, permits	Taxi Regulations	/
11	cloud, data, big_data, computing, customer, applications, digital_transformation, solutions, analytics, digital, technologies, blockchain, iot, devices, customers, systems, virtual, technology, processing, things	Cloud Computing	Digital Manufacturing

Table 7: Complete List of Topics and Topic Categories *(continued)*

Topic	Top FREX Words	Topic Label	Aggregated Topic Category
12	entrepreneurs, academics, investors, financing, entrepreneurship, loans, capital, credit, bank, start_ups, startups, loan, entrepreneur, growth, entrepreneurial, funds, venture_capital, academic, investor, startup	Entrepreneurship	/
13	transformation, technological, productive, strategies, competitive, textile, processes, value_chain, sector, sectors, efficiency, process, components, enablers, technologies, digitization, integration, competences, elaboration, structured	Transformation of the Production Process	Digital Manufacturing
14	determination, data_protection, flexibility, flexible, collective_agreements, collective_bargaining, co, working, interests, councils, seize, federal, social_partners, employers, unions, shaping, participation, options, working_conditions, statutory	Work Councils	/
15	year, turnover, quarter, sales, first, months, revenues, largest, last, half, annual, group, third, grow, forecast, total, forecasts, figures, compared, trend	Economic Statistics	/
16	digital_skills, skills, skill, employability, learning, graduates, literacy, training, cognitive, apprenticeship, solving, career, transversal, learn, mastery, apprenticeships, digital_technologies, careers, qualifications, soft	Digital Skills	Digital Investments
17	cars, car, electric, driving, driverless, roads, autonomous, bus, truck, vehicles, trucks, road, vehicle, parking, cities, drones, city, wheel, aviation, flight	Autonomous Vehicles	Automation & Compensation
18	theme, path, capable, born, excellence, explains, innovation, starting, chain, widespread, dedicated, destined, underlines, universities, shapes, choices, center, leap, thanks, hi	Other	/
19	robot, robots, robotic, robotics, humans, arm, arms, humanoid, programmed, hands, brain, beings, science_fiction, vacuum, human, 1950s, body, colleague, movements, moves	Robots	Automation & Compensation
20	profiles, degrees, artistic, occupation, animation, technicians, sound, competences, profile, audiovisual, graphic, content, programming, design, graduated, designer, visual, professionals, training, techniques	Creative Skills	Digital Investments
21	labour, independent, platform, knowledge_intensive, platforms, freelancers, clients, professionals, firms, earnings, freelance, transaction, might, tasks, firm, wage, skilled, temporary, consultants, organising	Independent Platform Work	Status Platform Workers
22	policies, productive, social_dialogue, administrations, competitiveness, reforms, employment, labor, essential, digitization, fabric, adequate, priority, youth, country, commitment, necessary, deficit, regulatory, reform	Competitivness Reforms	Digital Industrial Policy

Table 7: Complete List of Topics and Topic Categories *(continued)*

Topic	Top FREX Words	Topic Label	Aggregated Topic Category
23	talent, recruitment, managers, candidates, hr, executives, hiring, leadership, recruit, candidate, recruiting, surveyed, teams, respondents, leaders, believe, talents, professionals, management, selection	Human Resources	/
24	higher_education, college, education, university, courses, educational, students, student, vocational, colleges, universities, continuing_education, secondary, doctoral, institutions, programs, school, exam, lifelong_learning, grants	Higher Education	Digital Investments
25	uber, taxi_drivers, drivers, delivery, passengers, restaurants, protest, deliveries, riders, eat, driver, strike, taxis, couriers, app, bicycle, restaurant, passenger, rings, corporations	Gig Work	Status Platform Workers
26	graph, television, video, advertising, music, content, games, audiovisual, publishing, books, consumption, internet, tv, newspapers, turnover, game, billing, film, radio, distribution	Multimedia	/
27	think, thing, say, want, lot, going, look, talk, get, things, much, ca, feel, money, got, bad, tell, ask, happy, go	Other	/
28	jobs, automation, occupations, automated, workforce, skilled, productivity, replaced, wages, wave, inequality, job_losses, likely, rise, risk, advances, economist, roles, routine, economists	Automation	Automation & Compensation
29	sharing_economy, platforms, consumer, transactions, sharing, rental, collaborative_economy, users, rent, legal, platform, trader, law, liability, disputes, housing, dispute, rules, legislation, user	Sharing Economy	/
30	amortization, hyper, investments, incentives, super, depreciation, tax_credit, interventions, machinery, plan, continuation, incentive, maneuver, budget, decree, relief, subsidized, purchase, extension, credit	Investment Incentives	Digital Industrial Policy
31	century, humanity, lives, era, book, societies, capitalism, history, 21st_century, planet, 19th, intellectual, live, philosopher, revolutions, revolution, imagine, man, invention, let	Digital Future	/
32	agenda, member, regional, issues, reports, representatives, adopted, agendas, members, report, proposals, commission, consultation, county, forum, follow, dialogue, meetings, appointed, meeting	EU Agenda	/
33	labor_market, unemployment, proportion, unemployed, part_time, restructuring, educated, unions, structural, salary, increased, decreased, groups, figure, differs, wage, longer, union, extent, fixed	Labor Market Inequality	Automation & Compensation

Table 7: Complete List of Topics and Topic Categories *(continued)*

Topic	Top FREX Words	Topic Label	Aggregated Topic Category
34	smes, promote, projects, initiatives, tourism, actions, promotion, promoting, collaboration, facilitate, initiative, aimed, awareness, entities, implementation, aid, support, objectives, programs, lines	SME Support	Digital Industrial Policy
35	tech, said, chief_executive, voice, chief, centre, firms, might, officer, banks, firm, banking, bank, organisations, founder, seeing, centres, lawyers, biggest, office	Business News	/
36	broadband, infrastructure, connectivity, network, networks, telecommunications, energy, mobile, coverage, smart, operators, connected, deployment, fiber, high_speed, fast, connections, electricity, wireless, buildings	Digital Infrastructure	Digital Investments
37	municipalities, municipal, respondents, extent, individuals, assignments, survey, home, estimated, municipality, libraries, reduction, mail, costs, effect, net, estimate, expected, websites, differences	Municipalities	/
38	union, bargaining, relations, representation, autonomy, citizenship, inequalities, social, institutional, organizational, contractual, democratic, collective, participation, decent, reflection, organization, labor, forms, relationships	Industrial Relations	/
39	incidents, security, attacks, trust, cyber, privacy, expressed, page, answer, cybersecurity, suffered, continuity, attack, communications, micro, files, incident, consulted, damage, survey	Cyber Attacks & Data breaches	/
40	further_training, qualification, operational, employees, councils, mechanical_engineering, department, continuing_education, processes, requirements, learning, networking, topics, agile, participation, map, departments, digitization, involved, employee	Further Training	Digital Investments
41	programme, deliver, businesses, organisations, programmes, ensure, enable, enterprise, ensuring, delivering, adoption, local, wider, approach, address, supporting, engagement, digitalisation, range, vital	Business Support	Digital Industrial Policy
42	free, personal_data, users, user, search, neutrality, web, site, exploitation, sites, online, loyalty, advertising, consent, collection, audience, engine, freedom, blogs, terminals	Data Protection & Internet Regulation	/
43	professions, robotization, disappear, automation, routine, qualified, specialists, jobs, replace, tasks, threatened, original, replaced, job, labor_market, repetitive, profession, machines, experts, author	Automation	Automation & Compensation

Table 7: Complete List of Topics and Topic Categories *(continued)*

Topic	Top FREX Words	Topic Label	Aggregated Topic Category
44	gig_economy, minimum_wage, tribunal, gig, riders, rights, ruling, holiday_pay, protections, insecure, couriers, contractors, contracts, pay, sick, status, workers, worker, entitled, courier	Gig Worker Rights	Status Platform Workers
45	networked, topic, networking, medium_sized, boss, manufacturer, federal, wants, digitalization, mechanical_engineering, sees, politics, says, location, standards, board, corporations, association, enormous, head	Digital Manufacturing	Digital Manufacturing
46	plan, client, clients, brand, team, store, customer, best, marketing, product, ideas, purchase, experience, mind, channels, sell, brands, channel, commerce, moment	Business Strategy	/
47	exports, scenarios, crisis, elaborations, imports, diversification, ranking, economies, manufacturing, weight, goods, trade, foreign, export, added, specialization, geographical, relative, recorded, dynamics	Exports	/
48	president, edition, conference, yesterday, director, meeting, explained, event, organized, vice, held, stressed, deputy, attended, presentation, speakers, dedicated, headquarters, fair, head	High-level meetings	/
49	women, teachers, school, teaching, gender, schools, female, girls, men, teacher, students, children, science, male, parents, mathematics, classroom, boys, stereotypes, educational	Women & STEM	Digital Investments
50	research, efforts, r&d, universities, innovation, researchers, funding, colleges, entrepreneurship, collaboration, grants, climate, environments, scientific, institutes, strategic, institutions, societal, evaluations, grant	Research & Development	Digital Investments
51	factory, sensors, plant, assembly, maintenance, plants, components, factories, machine, warehouse, machines, additive_manufacturing, glasses, production, logistics, printing, parts, mechanical, manufacturers, augmented_reality	Smart Factory	Digital Manufacturing
52	self_employment, work_environment, self_employed, employer, false, investigation, hired, responsibility, client, contractor, employee, persons, work-, staffing, phenomenon, assignments, self-, employed, safety, contractors	(False) Self-Employment	/
53	election, politics, politicians, political, vote, party, left, elections, presidential, voters, reform, governments, minister, immigration, wing, chairman, liberal, reforms, anti, campaign	Elections	/
54	r&d, aid, clusters, expenditure, mission, creators, high_speed, favor, patent, fund, equity, venture_capital, heart, funding, patents, innovative, deployment, incubators, funds, seed	Industrial Policy	Digital Industrial Policy

Table 7: Complete List of Topics and Topic Categories *(continued)*

Topic	Top FREX Words	Topic Label	Aggregated Topic Category
55	competition, market, markets, economy, prices, price, global, consumers, profits, players, competitors, value, currency, traditional, margins, giants, monopoly, goods, consumption, profitable	Competition	/
56	vat, taxation, article, taxable, decree, invoice, obligations, paragraph, directive, invoicing, obligation, discipline, purposes, profits, tax, art, entities, electronic, establishment, compliance	Taxation	Taxation
57	approval, tax_system, approved, tax, fees, tax_evasion, crime, investigation, fraud, taxes, applicant, abuse, revocation, grounds, nutritional, chapter, deductions, conduct, error, section	Tax Evasion	Taxation
58	competence, digitalisation, efforts, society, opportunities, development, increased, small, needed, possibilities, contribute, welfare, goals, county, industries, important, parts, actors, businesses, regional	Digital Competences	Digital Investments
59	income, pension, insurance, retirement, contributions, basic, taxes, social_security, dividends, universal, compensation, unemployment, paid, self_employed, wages, unconditional, pay, salary, allowance, welfare	Compensation	Automation & Compensation
60	disappears, choices, productivity_gains, unit, evolutions, salaried, dematerialization, developments, poses, progress, emancipation, deregulation, divide, stagnation, tion, intervention, transforms, struggles, isolation, forms	Other	/

References

Blei, David M. 2012. "Probabilistic topic models." *Communications of the ACM* 55(4): 77.

Boukes, Mark et al. 2020. "What's the tone? Easy doesn't do it: Analyzing performance and agreement between off-the-shelf sentiment analysis tools." *Communication Methods and Measures* 14(2): 83–104.

Chang, Jonathan et al. 2009. "Reading tea leaves: How humans interpret topic models." *Advances in Neural Information Processing System*.

Gilardi, Fabrizio, Charles R. Shipan, and Bruno Wüest. 2020. "Policy diffusion: The issue–definition stage: American journal of political science." *American Journal of Political Science*.

Hu, Minqing, and Bing Liu. 2004. "Mining and summarizing customer reviews." *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*: 168–177.

Jockers, Matthew. 2017. "Package 'syuzhet'." http://mirrors.nics.utk.edu/cran/web/packages/syuzhet/syuzhet.pdf.

Mihalcea, Rada, and Paul Tarau. 2004. "TextRank: Bringing order into texts."

Mohammad, Saif M., and Peter D. Turney. 2010. "Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon." *CAAGET '10: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*: 26–34.

Moody, Christ. 2017. "Stop using word2vec." https://multithreaded.stitchfix.com/blog/2017/10/18/stop-using-word2vec/.

Nicholls, Tom, and Pepper D. Culpepper. 2020. "Computational identification of media frames: Strengths, weaknesses, and opportunities." *Political Communication*: 1–23.

Nielsen, Finn Årup. 2011. "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs." http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/6006/pdf/imm6006.pdf.

Quinn, Kevin M. et al. 2010. "How to analyze political attention with minimal assumptions and costs." *American Journal of Political Science* 54(1): 209–228.

Rinker, Tyler W. 2018. *lexicon: Lexicon data*. Buffalo, New York. http://github.com/trinker/lexicon.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2018. "Stm: R package for structural topic models." *Journal of Statistical Software* 10(2).

Rose, Stuart et al. 2010. "Automatic keyword extraction from individual documents." In *Text mining*, eds. Michael W. Berry and Jacob Kogan. Chichester, UK: John Wiley & Sons, Ltd, p. 1–20.

Silge, Julia. 2017. "Tidy word vectors, take 2!" https://juliasilge.com/blog/word-vectors-take-two/.

Spirling, Arthur, and Pedro L. Rodriguez. 2019. "Word embeddings what works, what doesn't, and how to tell the difference for applied research."

van Atteveldt, Wouter, Mariken A. C. G. van der Velden, and Mark Boukes. 2021. "The validity of sentiment analysis:comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms." *Communication Methods and Measures*: 1–20.

Wijffels, Jan. 2019. "Textrank." https://cran.r-project.org/web/packages/textrank/textrank.pdf.

Ying, Luwei, Jacob M. Montgomery, and Brandon M. Stewart. 2019. "Inferring concepts from topics: Towards procedures for validating topics as measures."

Young, Lori, and Stuart Soroka. 2012. "Affective news: The automated coding of sentiment in political texts." *Political Communication* 29(2): 205–231.