The diffusion of NLP methods in marketing research

A systematic analysis

Olivier Caron Christophe Benavent

May 6, 2024

Research context

NLP methods are increasingly used in marketing research

NLP methods enable the conversion of text into quantifiable data for in-depth analysis enabling the analysis of large volumes of text.

They are particularly suitable for marketing concerns:

- Evaluate customer feebacks, gauging public sentiment
- Detect emotional responses to products and marketing campaigns
- Detect trends, consumer preferences and needs.

Specific research practices

Institutional Framework promote competition in academia driven by journal rankings and citation scores, influencing researcher compensation and career advancement. (Richard et al., 2015)

Concentration of research in major publishers (Elsevier, Thomson) with an increasing number of journals with facilitated access through bibliographic search interfaces.

Research strategies

Researcher must elaborate strategies to advance their career (Kolesnikov et al., 2018)

- Balance between productivity, impact, originality, legitimacy, learning costs.

These strategies can be thought of in terms of the technology acceptance model from Davis (1989)

- Adopt NLP methods based on interest in innovative tools (1) and the perceived ease of integrating these methods into their work (2)

  –

1. Researchers are strategically adopting NLP methods to enhance their productivity and originality in marketing research
2. infusion of NLP methods in marketing research driven by researchers' interest in innovative tools and the perceived ease of integrating these methods into their work.
3. deliberate choice to balance research output (productivity) with the creation of impactful and original work within the marketing discipline.
4. Researchers might lean towards NLP methods that offer a low barrier to entry and high potential for enhancing the visibility and legitimacy of their work.

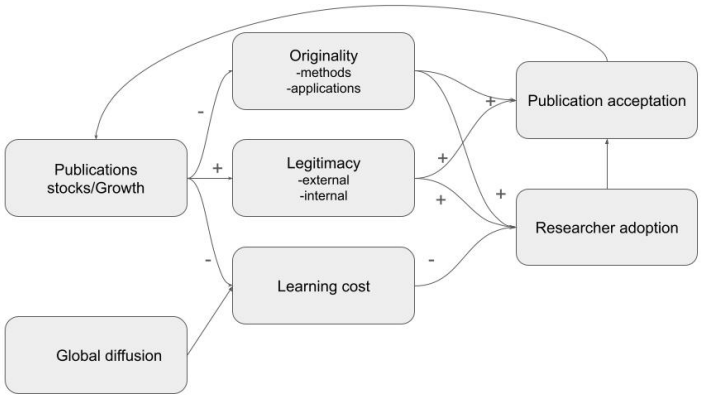
Management trends & communities

NLP methods are used in close communities of practice (Hauser et al., 2006; Lave & Wenger, 1991) where peer learning and shared practices enable the diffusion of methods

Management trends (Abrahamson, 1991), diffusion is linked to a cultural phenomenon rather than rational decision making

1. Community Practice: The adoption of NLP methods is fostered within tight-knit scholarly communities, suggesting that peer learning and shared practices play a crucial role in their diffusion.
2. Diffusion linked to CULTURAL PHENOMENON RATHER THAN RATIONAL DECISION MAKING

Conceptual framework



Data presentation

Articles

- There are 682 articles and 2310 different authors
- Date of publication range from 1985 to 2023

Data collection

- All data were collected from Scopus.

Keywords

- "natural language processing", "nlp", "embeddings", "chatgpt", "liwc", "transformers", "word2vec", "word-tovec", "lda", "text mining", "text-mining", "text analysis", "text analytics", "text-analytics", "text-analysis"

Journals

- 156 journals ranked from the first to third quartile according to the SCImago Journal Rank in the field of marketing.

General data vizualisation

Production per affiliation

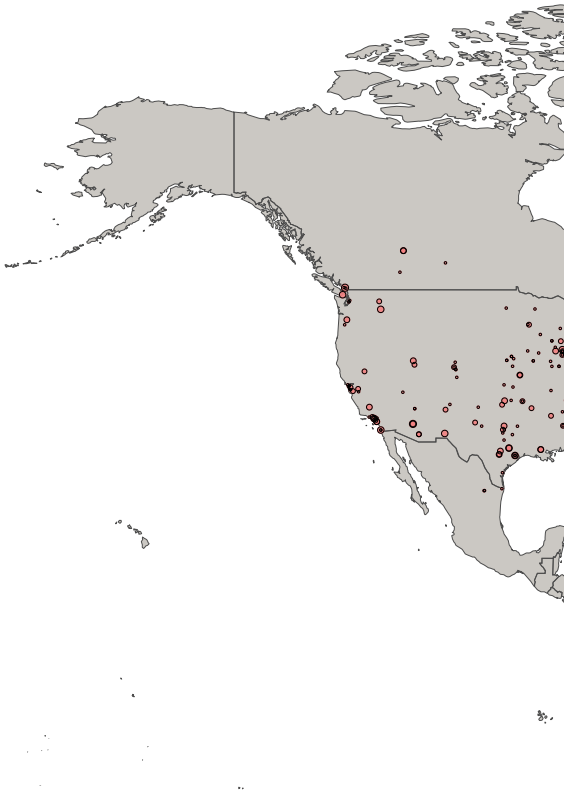
```
#have a look at countries on a worldmap
#specify some theme properties
plain <- theme(
  axis.text        = element_blank(),
  axis.line        = element_blank(),
  axis.ticks       = element_blank(),
  panel.border     = element_blank(),
  panel.grid       = element_blank(),
  axis.title       = element_blank(),
  panel.background = element_rect(fill = "white"),
  plot.title = element_text(hjust = 0.5)
)

#important to include the country name and the city name so that tidygeocoder can find the righ
#exemple : I had some problems with the city "Cambridge" which is in the US and in the UK. Now
productive_affiliations$city_country <- paste(productive_affiliations$affiliation_city, product

#enables us to get latitude and longitude of affiliation cities so we can place the cities on a
#It's a bit long so I commented it and saved the result in a csv file
#result_tidygeocoder <- tidygeocoder::geocode(productive_affiliations,
```

```
                                        #city_country,
                                        #method="osm") #OpenStreetMap data

#write_csv(result_tidygeocoder, "data/affiliations_geocoded.csv")

result_tidygeocoder <-read.csv("data/affiliations_geocoded.csv")

result_tidygeocoder <- result_tidygeocoder %>% filter(!is.na(affiliation_city))

mapWorld <- borders("world", colour="gray30", fill="#cbc8c3", size=0.2)
world <- map_data("world")
worldplot <- ggplot() +
  geom_polygon(data = world, aes(x = long, y = lat, group = group)) +
  geom_point(data = result_tidygeocoder, color="#f38181", alpha=0.9, aes(x=long, y = lat, size
  geom_point(data = result_tidygeocoder,
             colour="black",
             shape=1,
             aes(x=long, y = lat,
                 size = number_productions,
                 text= paste0("Affiliation: ",
                              affilname,
                              "\nAffiliation city: ",
                              affiliation_city,
                              "\nNumber of productions: ",
                              number_productions),
                 stroke = 0.20))+ #stroke is the width of the second black point and shape=1 m
  mapWorld+
  plain+
  coord_map("equirectangular")+
  coord_cartesian(ylim = c(-50, 90))+ #get rid of Antarctica;
  labs(title="")+
  scale_size(range = c(1, 6))
```

```
Warning in geom_point(data = result_tidygeocoder, colour = "black", shape = 1,
: Ignoring unknown aesthetics: text
```

```
Coordinate system already present. Adding new coordinate system, which will
replace the existing one.
```

```
worldplot_plotly_object <- ggplotly(worldplot, tooltip = "text")
worldplot_plotly_object %>%
  config(scrollZoom = TRUE)
```



```
htmlwidgets::saveWidget(worldplot_plotly_object, "images/worldplot.html")
```

Citations per country

A focus on affiliations: number of productions

A focus on affiliations: number of citations

Management trend in publication volume

```
nlp_papers <- list_articles

#get rid of conference papers
#nlp_papers_journal_only <- nlp_papers %>%
  #filter(!grepl("conference", subtypeDescription, ignore.case = TRUE)) #%>%
  #filter(year < 2023)


t0 <-prop.table(table(nlp_papers$`prism_publicationName`)) %>%
  as.data.frame() %>%
  arrange(desc(Freq)) %>%
  mutate(Var1 = sub(".*:", "", Var1)) %>% #just because title of Proceedings of the Academy...
  head(30)

t0 <- left_join(t0, scimago, by = c(Var1 = "Title"))

t1<-as.data.frame(table(nlp_papers$year))

g01 <- ggplot(t0, aes(x=reorder(paste(Var1, "-", SJR_Quartile, "| Rank:", Rank), Freq), y=Freq
  geom_bar(stat="identity", fill="steelblue") +
  coord_flip() +
  theme_minimal(base_size = 12) +
  labs(title="Number of Articles per Journal", y="Proportion", x="") +
  theme(
    axis.title.x = element_text(size = 16),
    axis.text.x = element_text(size = 14, angle = 45, hjust = 1),
    axis.text.y = element_text(size = 10),
    plot.title = element_text(size = 16)
  )
```

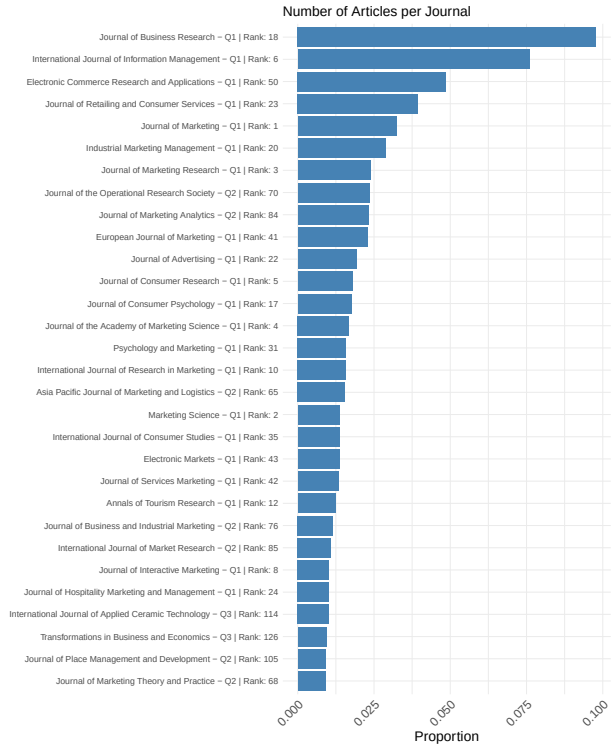
```
# Graph 2: Number of publications per year
g02 <- ggplot(t1, aes(x=Var1, y=Freq, group=1)) +
  geom_line(size=1.1, color="steelblue") +
  geom_point(size=2, color="steelblue") +
  geom_smooth(color="#7D7C7C", linewidth=0.5)+
  theme_minimal() +
  labs(title="Number of Publications per Year", y="", x="Year") +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size = 14),
    axis.title.x = element_blank(),
    plot.title = element_text(hjust = 0.5, size = 16) # Set title size to 16
  )
```

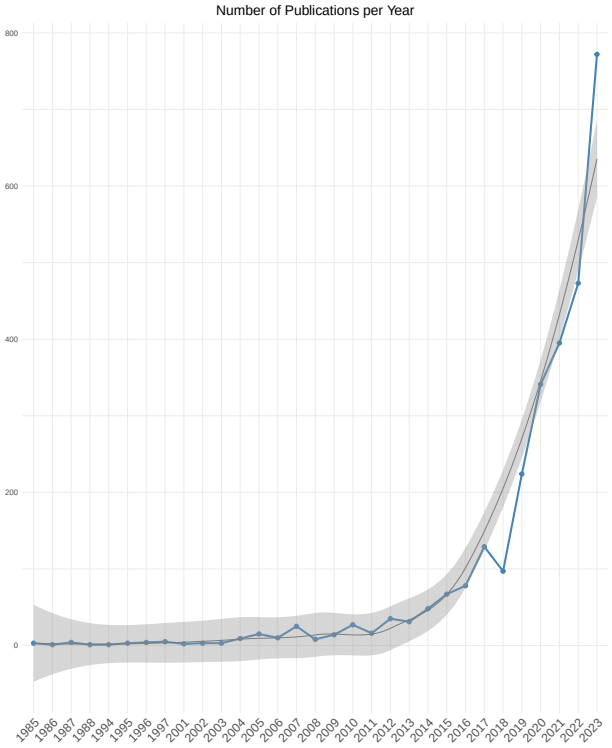
```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```

```
plotgrid <- plot_grid(g01,
                      g02,
                      label_size = 10,
                      ncol=2,
                      rel_widths =  c(1,1))
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

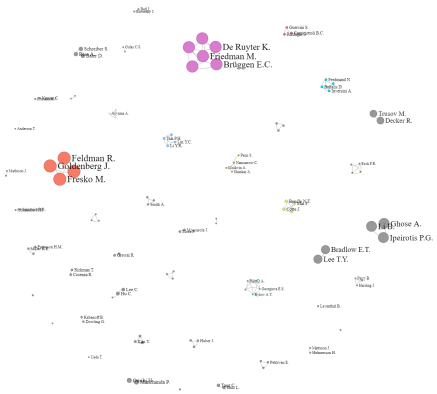
```
ggsave(filename="images/evolution_publications_nlp_marketing.png",
       width = 80,
       height = 40,
       units = "cm")
plotgrid
```



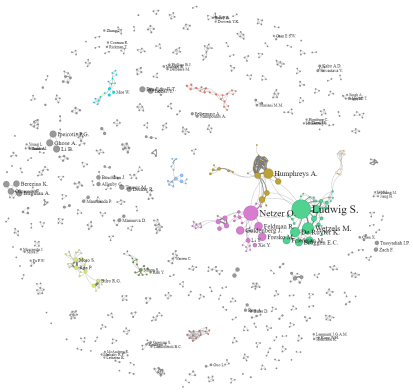


Networks & Communities

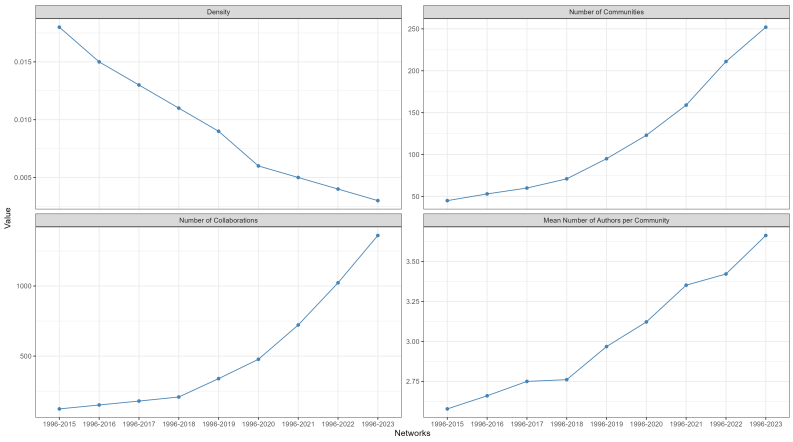
The structure of author networks (until 2015)



The structure of author networks (until 2023)



Some measures of network structure

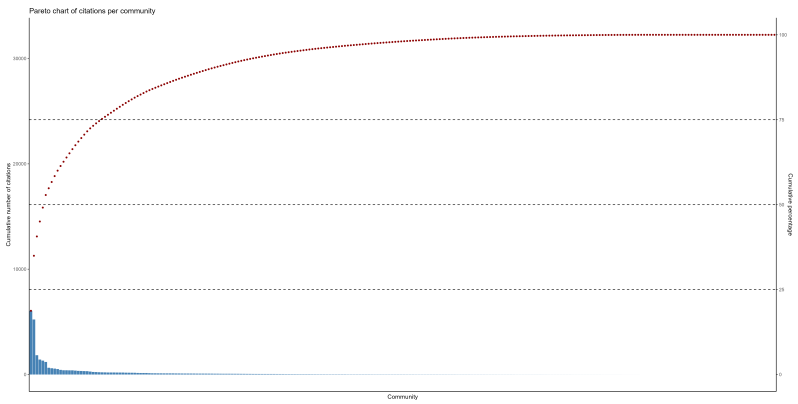


1. Network growth is outpacing the formation of new collaborations, leading to a less densely connected community.
2. The increasing number of communities reflects a diversification in research subjects and group formations.
3. A surge in collaborations suggests a trend towards more collective research efforts in NLP.

4. The expanding average size of research groups per community indicates the formation of larger, more collaborative teams.

The unequal spread of citations between communities

The top 5 communities account for 13.65% of authors (126) and 46% of citations (14851)



Topics

Topic modeling with STM

- STM algorithm with years as covariates (Roberts et al., 2014)

what we can understand from the result of these topics is that

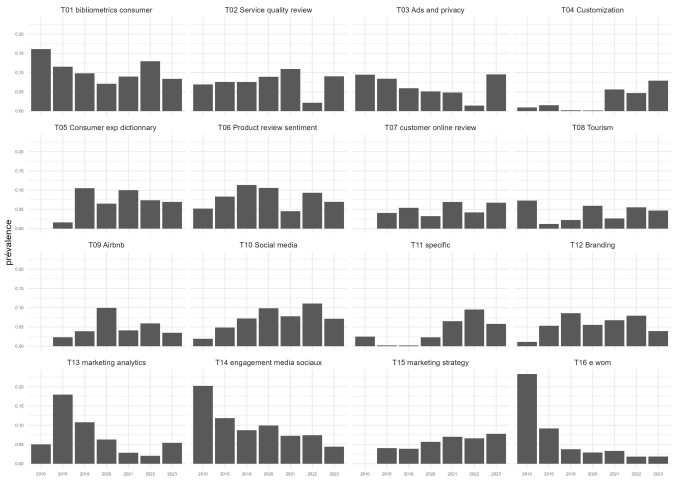
1. words like “customer”, “reviews” and “sentiment” : strong focus on consumer perceptions.
2. words related to service quality : improve service offerings.
3. terms like “brand,” “strategy,” and “engagement” appearing : strategic brand positioning.

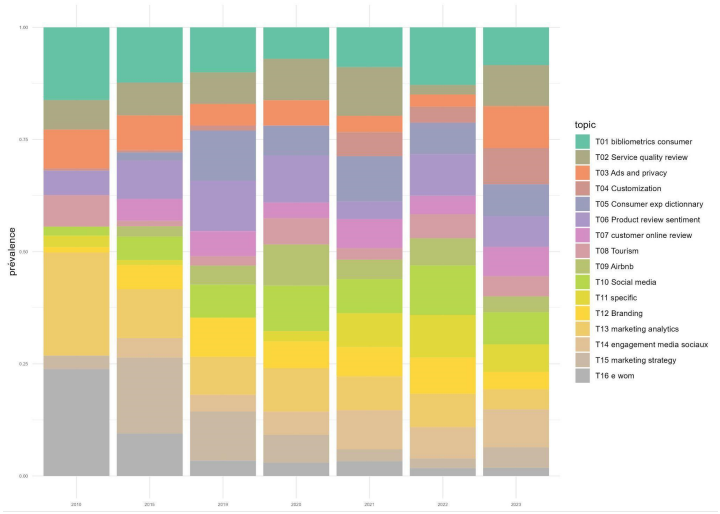
4. “social media” related terms indicates that platforms are key sources for data mining, providing rich insights for market trends and consumer behavior analysis.

5. “segmentation” with analytical terms shows an inclination towards using NLP for slicing the market into well-defined segments.

A concentric diffusion of topics

1. The number of topics per period increases steadily without the relative number of articles per topic decreasing, in a process of accumulation rather than substitution.





A focus on methods

Main NLP techniques used

- "O" or "1" if a technique is detected in the articles
- combined text of abstract, title, and keywords
- detection of the technique name and the different spellings

```r
bert_alt <- c(
  "bert",
  "bi-directional encoder representation from transformers",
  "bi-directional encoder representations from transformer",
  "bi-directional encoder representations from transformers",
  "bidirectional encoder representation from transformer",
  "bidirectional encoder representation from transformers",
  "bidirectional encoder representations from transformer",
  "bidirectional encoder representations from transformers",
  "bi directional encoder representation from transformer",
  "bi directional encoder representation from transformers",
  "bi directional encoder representations from transformer",
  "bi directional encoder representations from transformers"
)
```

```r
list_articles_techniques <- list_articles %>%
  distinct(entry_number, .keep_all = TRUE)

## NLP TECHNIQUES

liwc_alt <- c(
  "liwc",
  "linguistic inquiry and word count",
  "linguistic inquiry & word count",
  "linguistic inquiry word count",
  "linguistic word count"
)

chatgpt_alt <- c(
  "chatgpt", "chat gpts",
  "chat gpt", "chat gpts",
  "chat generative pre-trained transformer", "chat generative pre-trained transformers",
  "chat generative pretrained transformer", "chat generative pretrained transformers"
)
bert_alt <- c(
  "bert",
  "bi-directional encoder representation from transformer",
  "bi-directional encoder representation from transformers",
  "bi-directional encoder representations from transformer",
  "bi-directional encoder representations from transformers",
```

```
  "bidirectional encoder representation from transformer",
  "bidirectional encoder representation from transformers",
  "bidirectional encoder representations from transformer",
  "bidirectional encoder representations from transformers",
  "bi directional encoder representation from transformer",
  "bi directional encoder representation from transformers",
  "bi directional encoder representations from transformer",
  "bi directional encoder representations from transformers"
)

llm_alt <- c(
  "large language model", "large language models",
  "large scale language model", "large scale language models"
)

lda_alt <- c(
  "latent dirichlet allocation", "latent dirichlet allocations",
  "lda", "ldas",
  "dirichlet allocation", "dirichlet allocations"
)

stm_alt <- c(
  "structural topic model", "structural topic models",
  "structural topic modeling", "structural topic modelings",
  "structural topic modelling", "structural topic modellings"
)

ner_alt <- c(
  "named entity recognition", "named entity recognitions",
  "entity recognition", "entity recognitions"
)

sentiment_analysis_alt <- c(
  "sentiment analysis"
)

topic_modeling_alt <- c(
  "topic modeling", "topic modelings",
  "topic model", "topic models",
  "topic modelling", "topic modellings"
```

```
)

tfidf_alt <- c(
  "tf-idf", "tf-idfs",
  "tfidf", "tfidfs",
  "term frequency-inverse document frequency", "term frequency-inverse document frequencies",
  "term frequency inverse document frequency", "term frequency inverse document frequencies"
)

embeddings_alt <- c(
  "word embedding", "word embeddings",
  "embeddings"
)

transformers_alt <- c(
  "transformers",
  "transformers model", "transformer models"
)

roberta_alt <- c(
  "roberta", "robertas",
  "robustly optimized bert approach", "robustly optimized bert approaches"
)

word2vec_alt <- c(
  "word2vec", "word2vecs",
  "word to vec", "word to vecs",
  "word 2 vec", "word 2 vecs",
  "word2vec model", "word2vec models",
  "word to vectors", "word to vector",
  "word 2 vectors", "word 2 vector"
)

fasttext_alt <- c(
  "fasttext", "fasttexts"
)

textrank_alt <- c(
  "textrank", "textranks"
)
```

```
gpt2_alt <- c(
  "gpt2", "gpt-2", "gpt 2",
  "generative pre-trained transformer 2", "generative pre-trained transformers 2",
  "generative pretrained transformer 2", "generative pretrained transformers 2"
)

gpt3_alt <- c(
  "gpt3", "gpt-3", "gpt 3",
  "generative pre-trained transformer 3", "generative pre-trained transformers 3",
  "generative pretrained transformer 3", "generative pretrained transformers 3"
)

pos_tagging_alt <- c(
  "pos tagging", "pos taggings",
  "part of speech tagging", "parts of speech tagging",
  "pos tagger", "pos taggers"
)

para_alt <- c(
  "paralanguage classifier", "paralanguage classifiers",
  "textual paralanguage classifier", "textual paralanguage classifiers"
)

leximancer_alt <- c(
  "leximancer"
)

passivepy_alt <- c(
  "passivepy"
)

neural_networks_alt <- c(
  "neural network", "neural networks",
  "neural net", "neural nets"
)

transformers_alt <- c(
  "transformers",
  "transformers model", "transformer models"
)
```

```r
glove_alt <- c("glove", "global vectors for word representation", "global vectors")

t5_alt <- c("text-to-text transfer transformer", "text-to-text transfer transformers", "text to
camembert_alt <- c("camembert", "camemberts")
flaubert_alt <- c("flaubert", "flauberts")

# count the presence of each technique in the abstracts, title, and keywords
check_presence <- function(text, keywords) {
  sapply(keywords, function(keyword) grepl(keyword, text, ignore.case = TRUE)) %>% any()
}

# Add columns for each technique based on above dictionaries
techniques <- list(LIWC = liwc_alt, ChatGPT = chatgpt_alt, LLM = llm_alt, LDA = lda_alt, STM =
                   BERT = bert_alt, NER = ner_alt, Sentiment_Analysis = sentiment_analysis_alt
                   Neural_networks = neural_networks_alt, Transformers = transformers_alt, Glo
                   TFIDF = tfidf_alt, Embeddings = embeddings_alt, Transformers = transformers_
                   Flaubert = flaubert_alt,
                   RoBERTa = roberta_alt, Word2Vec = word2vec_alt, FastText = fasttext_alt,
                   TextRank = textrank_alt, GPT2 = gpt2_alt, GPT3 = gpt3_alt, POS_Tagging = pos
                   PARA = para_alt, Leximancer = leximancer_alt, PassivePy = passivepy_alt)

# create a column for each technique to which we apply the check_presence function to populate
for (technique_name in names(techniques)) {
  list_articles_techniques[[technique_name]] <- sapply(list_articles_techniques$combined_text,
  list_articles_techniques[[technique_name]] <- as.integer(list_articles_techniques[[technique_
}

# number of techniques per year
techniques_per_year <- list_articles_techniques %>%
  group_by(year) %>%
  summarise(across(all_of(names(techniques)), sum, na.rm = TRUE)) %>%
  ungroup()
```

```
Warning: There was 1 warning in `summarise()`.
i In argument: `across(all_of(names(techniques)), sum, na.rm = TRUE)`.
i In group 1: `year = 1985`.
Caused by warning:
! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
Supply arguments directly to `.fns` through an anonymous function instead.
```

```
  # Previously
  across(a:b, mean, na.rm = TRUE)

  # Now
  across(a:b, \(x) mean(x, na.rm = TRUE))
```

```
# cumulative number of techniques per year
cumulative_techniques_year <- techniques_per_year %>%
  arrange(year) %>%
  mutate(across(-year, cumsum))

techniques_2023 <- cumulative_techniques_year %>%
  filter(year == 2023) %>%
  pivot_longer(-year, names_to = "technique", values_to = "count") %>%
  select(-year) %>%
  arrange(desc(count))

top_10_techniques <- techniques_2023$technique  %>%
  head(10)

# Sum the number of techniques used in each article
list_articles_techniques$sum_techniques <- rowSums(list_articles_techniques[names(techniques)])
most_technical_articles <- list_articles_techniques[order(list_articles_techniques$sum_techniqu
  select(entry_number, dc_title, 23, year, sum_techniques, Rank, SJR, citedby_count)
```

Evolution of NLP techniques

```
techniques_2023 %>%
  pivot_wider(names_from = technique, values_from = count) %>%
  gt() %>%
  tab_options(
    column_labels.font.weight = "bold"
  )
```

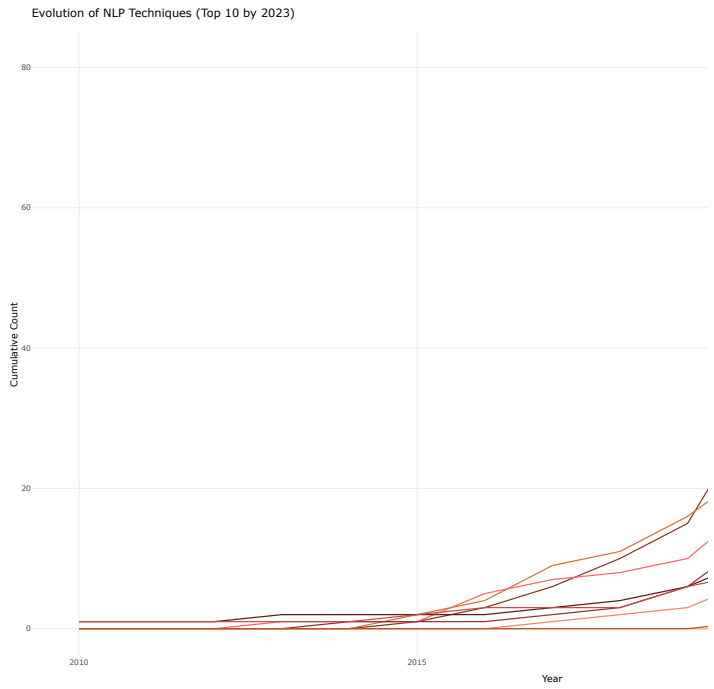
Sentiment_Analysis	Topic_Modeling	LDA	LIWC	Neural_networks	ChatGPT	Embeddings	Lexima
81	68	53	26	26	20	19	

```r
#other technique to plot with label on curves

Royal2 <- wesanderson::wes_palette("GrandBudapest1", n = 10, type = "continuous")

fig2 <- ggplot(cumulative_techniques_year, aes(x = year)) +
  scale_color_manual(values = Royal2) +
  xlim(2010, 2024) +
  labs(
    title = "Evolution of NLP Techniques (Top 10 by 2023)",
    subtitle = "Cumulative sum of articles mentioning each technique",
    x = "Year",
    y = "Cumulative Count"
  ) +
  theme_minimal() +
  theme(legend.position = "none") # hide legends since they are on the curves

# Add lines for each technique
for (technique in top_10_techniques) {
  fig2 <- fig2 + geom_line(aes_string(y = technique, color = sprintf("'%s'", technique)))
}
```

```
Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
i Please use tidy evaluation idioms with `aes()`.
i See also `vignette("ggplot2-in-packages")` for more information.
```

```r
data_2023 <- cumulative_techniques_year[cumulative_techniques_year$year == 2023, ]


labels_positions <- data.frame(year = 2023)

for (technique in top_10_techniques) {
  if (technique %in% names(data_2023)) {
    labels_positions[[technique]] <- data_2023[[technique]]
  }
}

fig_plotly <- fig2 #to use without geom_text_repel but geom_text because not supported yet by p
for(technique in names(labels_positions)[-1]) {
  fig_plotly <- fig_plotly + geom_text(data = labels_positions,
                                aes_string(x = "year", y = technique, label = gsub("_", " ", spr
```

```
                                  nudge_x = 0.5, nudge_y = 0)
}
ggplotly(fig_plotly)
```



Diffusion and delay of NLP techniques

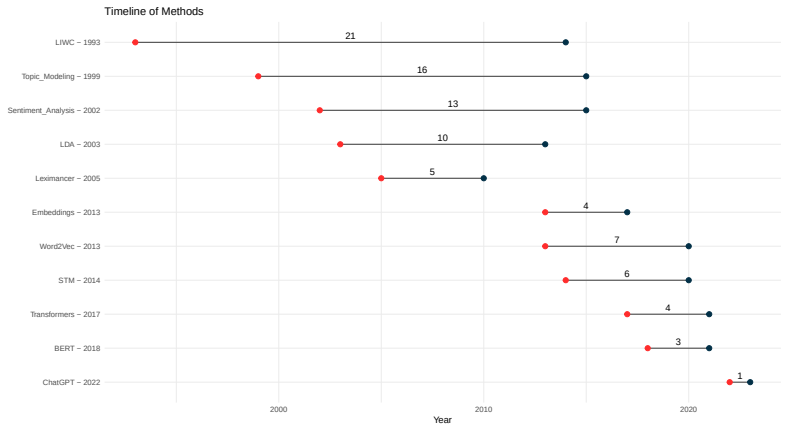
```
first_use <- data.frame(technique = character(), year = integer())

# Loop through each column starting from the 2nd column
```

```r
for (col in 2:ncol(techniques_per_year)) {
  # Find the first row where the value is greater than 0
  first_row <- which(techniques_per_year[[col]] > 0)[1]

  # Check if there is at least one value greater than 0
  if (length(first_row) > 0) {
    # Get the year from the first column for this row
    year <- techniques_per_year$year[first_row]

    # Append to the first_use dataframe
    first_use <- rbind(first_use, data.frame(technique = names(techniques_per_year)[col], year_
  }
}

first_use <- first_use %>%
  arrange(desc(year_first_use))
```

```r
#NER source : Nadeau, David & Sekine, Satoshi. (2007). A Survey of Named Entity Recognition and

first_use$year_publication[first_use$technique == "ChatGPT"] <- "2022"
first_use$year_publication[first_use$technique == "LIWC"] <- "1993"
first_use$year_publication[first_use$technique == "VADER"] <- "2004" #Hutto, C.J. & Gilbert, E
#first_use$year_publication[first_use$technique == "LLM"] <- "2022"
first_use$year_publication[first_use$technique == "LDA"] <- "2003"
first_use$year_publication[first_use$technique == "STM"] <- "2014"
first_use$year_publication[first_use$technique == "BERT"] <- "2018"
first_use$year_publication[first_use$technique == "NER"] <- "1996" #Sixth Message Understanding
first_use$year_publication[first_use$technique == "Sentiment_Analysis"] <- "2002" #Pang B, Lee
first_use$year_publication[first_use$technique == "Topic_Modeling"] <- "1999" #Hofmann, T. (199
first_use$year_publication[first_use$technique == "TFIDF"] <- ""
first_use$year_publication[first_use$technique == "Embeddings"] <- "2013"
first_use$year_publication[first_use$technique == "Transformers"] <- "2017"
#first_use$year_publication[first_use$technique == "RoBERTa"] <- "2019"
first_use$year_publication[first_use$technique == "Word2Vec"] <- "2013"
first_use$year_publication[first_use$technique == "FastText"] <- "2015"
first_use$year_publication[first_use$technique == "TextRank"] <- "2004"
#first_use$year_publication[first_use$technique == "GPT2"] <- "2019"
first_use$year_publication[first_use$technique == "GPT3"] <- "2020"
#first_use$year_publication[first_use$technique == "POS_Tagging"] <- ""
#first_use$year_publication[first_use$technique == "PARA"] <- "" ## 2023 but it's the paper tha
```

```r
first_use$year_publication[first_use$technique == "Leximancer"] <- "2005"
#first_use$year_publication[first_use$technique == "PassivePy"] <- "" ## 2023 but same as PARA
#first_use$year_publication[first_use$technique == "LDA2Vec"] <- "2016"
#first_use$year_publication[first_use$technique == "LSTM"] <- "1997"
#first_use$year_publication[first_use$technique == "GRU"] <- "2014"
#first_use$year_publication[first_use$technique == "RNN"] <- "1986"
#first_use$year_publication[first_use$technique == "CNN"] <- "1989"



first_use <- first_use %>%
  filter(year_publication != "" & year_first_use != "") %>%
  mutate(year_publication = as.integer(year_publication),
         year_first_use = as.integer(year_first_use),
         delay  = year_first_use - year_publication,
         technique_year = paste(technique,"-",year_publication)) %>%
  arrange(year_publication)

## I only change the order of the rows to make the graph more readable. I want embeddings to a

# Identify the index of the rows
index_embeddings <- which(first_use$technique == "Embeddings")
index_word2vec <- which(first_use$technique == "Word2Vec")

# SAve the rows
ligne_embeddings <- first_use[index_embeddings, ]
ligne_word2vec <- first_use[index_word2vec, ]

# Switch the rows
first_use[index_embeddings, ] <- ligne_word2vec
first_use[index_word2vec, ] <- ligne_embeddings

techniques_lolipop <- ggplot(first_use, aes(x = reorder(technique_year, year_publication), y =
  geom_segment(aes(xend = technique_year, yend = year_publication), colour = "#3a3a3a", linewi
  geom_point(aes(y = year_first_use), colour = "#023047", size = 2.5) +
  geom_point(aes(y = year_publication), colour = "#ff2c2c", size = 2.5) +
  geom_text(aes(x = technique_year, y = (year_first_use + year_publication) / 2, label = delay
  coord_flip() +
  labs(y = "Year", x = "", title = "Timeline of Methods") +
  scale_x_discrete(limits = rev(first_use$technique_year)) +
```

```
  theme_minimal()



techniques_lolipop
```




```
#ggplotly(techniques_lolipop, tooltip = c("technique", "year_publication", "year_first_use", "

ggsave("images/first_use_techniques.png", width = 10, height = 6, dpi = 600)
```

There is a significant time lag between the emergence of techniques and their actual use in marketing research. This delay is decreasing over time, suggesting that the cost of learning NLP is decreasing.

Dominantly marketing: the selective borrowing of NLP methods

Future of NLP in marketing

- LLMs lower learning costs for sentiment analysis and classification tasks
- LLMs enable agent based
- Fine-tuned machine learning methods still perform better for specific tasks (Krugmann & Hartmann, 2024)
- but LLMs achieve very good results without prior training

To conclude:

1. NLP methods' adoption in top-tier marketing journals suggests it's not a mere trend but an innovation gaining solid ground in the field.
2. The cost of learning NLP is decreasing over time, facilitated by a growing repository of shared knowledge.
3. The integration and diffusion of NLP techniques are bolstered within communities of practice.
4. Diffusion is driven by new research questions, data availability, and suitable techniques.
5. Delay between the introduction of a technique and its adoption is decreasing over time.

References

Abrahamson, E. (1991). Managerial fads and fashions: The diffusion and rejection of innovations. *The Academy of Management Review*, *16*(3), 586. https://doi.org/10.2307/258919

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319. https://doi.org/10.2307/249008

Hauser, J., Tellis, G. J., & Griffin, A. (2006). Research on Innovation: A Review and Agenda for "Marketing Science". *Marketing Science*, *25*(6,), 687–717. http://www.jstor.org/stable/40057216

Kolesnikov, S., Fukumoto, E., & Bozeman, B. (2018). Researchers' risk-smoothing publication strategies: Is productivity the enemy of impact? *Scientometrics*, *116*(3), 1995–2017. https://doi.org/10.1007/s11192-018-2793-8

Krugmann, J. O., & Hartmann, J. (2024). Sentiment Analysis in the Age of Generative AI. *Customer Needs and Solutions*, *11*(1), 3. https://doi.org/10.1007/s40547-024-00143-4

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511815355

Richard, J., Plimmer, G., Fam, K.-S., & Campbell, C. (2015). Publishing success of marketing academics: antecedents and

outcomes. *European Journal of Marketing*, *49*(1/2), 123–145. https://doi.org/10.1108/EJM-06-2013-0311

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, *58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Appendix