

Is this a Mash-Up or a Template? Towards a Typology for Literary Bots

Isabelle Gribomont^{1,2} and Lucie Mentalechta¹

¹UCLouvain

²KBR (Royal Library of Belgium)

This paper introduces the LabEL literary bot typology, which was developed empirically from 588 literary bot accounts on Twitter. The typology features eight categories that capture various aspects of bot procedural-ity. It abstracts the processes reliant on specific files, resources, tools, and scripts into stable elements. The typology is introduced with a contextual overview of the Twitter platform, since bots’ formal elements are influenced by their technosocial environment. We provide several case studies to illustrate the practical application of this typology, showcasing its strengths and limitations. Finally, we briefly explore potential applications for studying and preserving literary bots. In addition to offering tools for analysing literary bots individually, within collections, or alongside other types of generative electronic literature, the typology also aids in preserving a fragmented memory of this form and contextualising it within the broader landscape of digital literary art.

Keywords: Digital Literature, Literary Bots, Typology, Twitter, Twitterature

1 Introduction

Since the early days of computing, technical innovations have prompted literary experiments. From Christopher Strachey’s 1953 automatically generated love letters to poems written by ChatGPT, a history of digital literature would include all major advances in computing and Natural Language Processing. The development of hypertext technology, the World Wide Web and probability-based language models, to name a few, have all played a major role in the multifaceted history of digital literature. The combination of strict constraints and possibilities to transcend norms imposed by print media has provided a fertile ground for writers to innovate.

With the democratisation of informatics, especially the apparition of content publishing platforms in the late 1990s, digital literature has become increasingly accessible to laypersons. Since then, specialist knowledge is no longer a prerequisite to creatively engage with digital technologies. Although literary creation online still requires a degree of material and societal privilege, the advent of social media platforms has

further lowered the barrier to entry and led to the emergence of countless literary voices online.

Twitter is one such corner of the internet which has been quietly conquered by professional and amateur writers alike. The character limit becomes a structural constraint akin to a modern-day versification while the affordances of hashtags, threads, tags, bots, community access and a massive mode of distribution can be harnessed for aesthetic and conceptual innovations.¹ More than these new constraints and affordances, the specific context of a social media platform such as Twitter drastically alters the literary and publishing paradigms found in other contexts.

Literary bots, in particular, have been a popular literary form on Twitter due to the possibilities afforded by the API and services such as *Cheap Bot, Done Quick!* which allowed users to create bots without coding skills or an API account. A Twitter bot is a Twitter account which posts automatically, in reaction to other tweets or on a schedule. Misinformation and propaganda bots have received the most attention, but Twitter bots have been created with a wide range of objectives, including art, literature, activism and entertainment (Ortega, 2020). Like all literature on Twitter, literary bots belong to the third wave of electronic literature, as defined in Flores (2021b). This third wave is characterised by the use of digital platforms (i.e. content creation platforms, social media platforms and digital writing platforms), where aesthetic innovations are often replaced with "remixes" of existing content (Flores, 2021b). In contrast, digital works belonging to the first and second generations tend to prioritise formal and aesthetic experiments. This preoccupation is often contextualised within the literary and artistic traditions of avant-garde movements such as dadaism, surrealism and Oulipo. Nevertheless, literary Twitter bots tend to be more connected to the traditions of these avant-gardes than most other forms belonging to this third wave, to the point that it could be argued that they create a bridge between them.

The study of *Twitterature*, whether it is automatically generated or not, has been hindered by the difficulty of identifying and collecting literary tweets on a large scale. The research presented here is part of a larger initiative aiming to (semi-)automatically identify and label literary tweets in existing Twitter archives such as the general Twitter Stream grabbed by the Archive Team, which contains 6.7 TB of Twitter data streamed from 2012 to 2021.² This project would allow for the creation of a literary Twitter archive and, as a consequence, the emergence of more systematic studies of *Twitterature*. With the change in leadership and API solutions proposed, the future of literature on Twitter in general, and bots in particular, is uncertain. As the community started developing in other platforms, such as Mastodon, it is a good time to look back at the history of this form.

This paper introduces the LABEL literary bot typology, which was developed empirically from a corpus of 588 literary Twitter bot accounts.³ The outline of the typology is prefaced by a contextual overview of the Twitter platform and existing typologies. We then present a few case studies to demonstrate how this typology applies to practical cases and showcase its strength as well as its limitations. Finally, we briefly discuss possible applications in the context of the study and preservation of literary bots.

¹ Although Twitter was rebranded as "X" in July 2023, we have opted to use "Twitter" throughout this paper since it was the name in use during most of the relevant time period.

² URL: <https://archive.org/details/twitterstream>

³ This research was developed in the context of LABEL (Laboratory for Electronic Literature), a collaborative research project at KBR (Royal Library of Belgium) and UCLouvain dedicated to digital literature.

2 Twitter and Twitter bots: a retrospective

Although the purpose of this article is not to analyse a corpus of literary Twitter bots, but to introduce the typology which emerged from it, we incidentally provide an overview of this highly media-specific form. Therefore, in what is likely to become a retrospective, this article is also an opportunity to offer preliminary reflections on the characteristics of literary Twitter bots, within their "technosocial context" (Marino, 2020). It is evident that the history of literary bots is intertwined with the history of the platform itself. The following overview focuses on the ways in which changes in the platform's leaderships and user base influenced the development of literary Twitter bots. Because Twitter is not primarily a literary platform and became a stage for high-stake political and ideological struggles, the increasing hostility of the Twitter environment for literary bots can be read as a collateral damage in a fight against disinformation and political propaganda. However, the reality is more complex, since the history of the platform is also intertwined with a struggle for profit and influence emanating from its leadership, where the disregard for not-for-profit, small-scale third-party initiatives should also be read politically.

2.1 The creation of a network

Twitter was created by Jack Dorsey, Biz Stone, and Evan Williams and launched in July 2006. It initially aimed to be a lightweight microblogging service that allowed users to share updates within a 140-character limit. It was imagined as a means to revolutionise communication on an interpersonal level. Twitter's initial concept revolved around a "status" service concept accessible via a website and via SMS in the United States. The character limit imposed by SMS constraints fostered creativity in crafting succinct messages and encouraged immediate interactions on the platform (Burgess and Baym, 2020). The minimalist interface and only guideline ("What's your status?") did not provide much clarity in the way it was to be used. In the words of Burgess and Baym: "Twitter's ambiguity almost demanded that its users develop their own ideas about what to do with it" (p. 7).

In its early days, public opinion was instrumental in shaping the platform. Features such as @ mentions, hashtags and retweets were suggested by users and incorporated in 2006 and 2007. If the implementation of such requests demonstrated an acknowledgment of the users' needs, it often resulted in new issues. For instance, the use of the hashtags to create trending topics became problematic, as users were encouraged to use the feature maliciously, to disseminate spam or engage in propaganda (Burgess and Baym, 2020).

2.2 From social network to (dis)information outlet

Twitter has undergone a significant evolution from a startup to a key player in the social media landscape. This evolution is the result of the multiple changes in governance, among other factors (Chang et al., 2023; Novoa et al., 2022). In 2009, Evan Williams, one of the founders and CEOs, declared: "Twitter is not a social network, it's an information network" (Burgess and Baym, 2020, p. 13). At this time, the platform's prompt changed from "What are you doing?" to "What's happening?". Williams took part in the construction of a world-centred, public and news-oriented Twitter (Burgess and Baym, 2020). The platform started to be used as an information channel, and its unique position at the intersection between a social network and a news

platform transformed it into a major stage of political activism, news dissemination, and international relations. For instance, Twitter was used extensively by Barack Obama during the 2008 presidential campaign. Since this particular event, such practices blending politics and social networks have become common place (Evans et al., 2014).

Its growing role in breaking news and its open door to citizen journalism have also led Twitter to become a “vector for disinformation”. The microblogging service is particularly susceptible to these assaults because of the casual form of the exchanges and asymmetrical structure of the network’s nodes: following a user does not grant any reciprocal effect, and the information flows only from the followed account to the follower. In addition, Twitter provides an opportunity for organisations that would otherwise not have had the resources to conduct (dis)information campaigns (Chamberlain, 2010).

Twitter bots, in particular, have played a major role in several key disinformation campaigns such as the 2016 elections in Europe and the US (Schiffrin, 2017) or post-truth spreading in Mexico with the “Peñabots” posting pro-Enrique Peña Nieto propaganda (Velasquez, 2019). More recently, during the COVID-19 pandemic, the company was forced to expand its policy against malicious uses of the platform, and its automated moderation tools have challenged over 1.5 million accounts for spam or manipulative behaviour in COVID-19 discussions (Singh and Blase, 2020). This aspect of the platform has brought negative press over the years and pushed the network to take measures aiming at limiting this phenomenon, notably by repeatedly updating their API permissions and content moderation policies.

2.3 Literary Twitter bots in the face of API changes, automated content moderation and Elon Musk

From 2008, to counter misinformation, but also to maintain a coherent brand image and protect users’ data, Twitter made iterative changes to its Application Programming Interfaces (APIs) and development policies. The company bought third-party applications deemed fit for incorporation, such as their search engine, from a company named Summize in 2008, and blocked others through API modifications. This strategy impacted mobile app developers and many of the services and tools initially developed independently from Twitter, such as URL shortening services and image-sharing applications. Towards 2010 and 2011, Twitter definitely turned away from “open innovation”, and marched toward a centralised and advertising-driven paradigm. From 2010, as the APIs matured, more restrictions were introduced, affecting third-party platforms such as Google, YouTube and Facebook (Burgess and Baym, 2020; Jünger, 2021).

Three types of access restriction were progressively incorporated into Twitter’s APIs. The first consisted in the introduction of open authorisation for API access (OAuth), requiring all requests to be authorised, marking a departure from the previous more open access policy. This first measure caused many developers to struggle with its implementation as well as the changes in user experience that came with it. This event is known on Twitter as the “#oauthcalypse”. The second mandated that third-party apps on mobile and entertainment devices, especially those with more than 100,000 users, undergo app reviews before they could operate in production mode, effectively limiting the development of new clients without Twitter’s collaboration. The third included the imposition of rate limits on all API endpoints, limiting the amount of data that could be requested within specific sliding time frames. These changes made access more complex and generated some negative reactions within the ecosystem,

shaping the foundation of Twitter's API as it stands today (Jünger, 2021; van der Mersch, 2016). The community of bots and apps thriving on Twitter before this shift was decimated by these changes, since users which were not in capacity to adapt to updates were disabled. As a result, only very few literary and artistic bots created before 2010 continued to operate (Flores, 2021a).

Following Elon Musk's acquisition of Twitter for 44 billion dollars in October 2022, drastic changes to the APIs have been implemented. At the time of writing, a year after the change in leadership, four different tiers are offered, as per the documentation published on the Twitter developer platform. The free tier allows 1,500 tweets to be posted per month, but does not allow any tweets to be read. The basic tier, priced as \$100/month, offers the option of posting 3,000 tweets per month, with a 10,000 tweets read limit. The pro tier, priced at \$5,000/month, offers a 300,000 tweets posting limit and a 1 million tweets reading limit. In addition, the pro tier offers access to the Filtered stream API, full-archive search, and the Ads API. Finally, the enterprise tier starts at \$42,000 per month and offers custom commercial-level access. Although some literary bots could technically operate using the free tier, those that use Twitter as their source would need the pro tier to function.⁴ Importantly, along with countless third-party applications, tools like *Cheap Bot*, *Done Quick!* which allowed users to create bots without coding skills and a Twitter developer account can no longer afford to operate after this latest change. On the 6th of April 2023, when the service closed down, a large portion of literary and artistic bots stopped posting.

In addition to API changes, automated content moderation is another long-standing enemy of literary bots. Around 2020, Twitter estimated that roughly 5% of its accounts were spam bots. The platform permits automated bots if they do not violate their terms. However, the company disclosed that it removed approximately one million spam bot accounts daily and temporarily locked millions more each week until their operators pass anti-spam tests. Over the years, Twitter increasingly relied on automated and machine learning content moderation tools, acknowledging that this may lead to moderation errors, including mistakenly suspending legitimate accounts, as the platform seeks to combat disinformation and malicious bots (Arroyo-Machado et al., 2023; Singh and Blase, 2020). The fight against spam bots has allegedly become more virulent under Musk, who has claimed that their prevalence reached 33% of users (Arroyo-Machado et al., 2023).⁵

Finally, the figure of Elon Musk has also contributed to driving literary and artistic bots away from the platform. When Musk bought Twitter, an unusual spike in accounts deactivation was noted (Chang et al., 2023). One of the automated tools tracking "inauthentic behaviour" on Twitter, *Bot Sentinel*, reported the deactivation of around 877,000 accounts and a further suspension of 497,000 accounts between October 27th and November 1st, 2022. These numbers hint at a 208% increase in account losses in the days after Musk purchased the company. Two factors are considered to be a possible explanation for these numbers. First, the deliberate account deletion could be the result of a movement of protest from the users, unhappy with the platform's integrity values with Musk at its head. Secondly, the raise in automated suspension of accounts could be the consequences of voluntary violations of Twitter's rules as a mean to test the new regulations under Musk's leadership (Stokel-Walker, 2022).

⁴ The academic API, which allowed researchers to harvest up to 10 million tweets per month is now defunct and no other solution has been offered to date.

⁵ Previous studies suggest different estimates, between 3% and 15% (Chavoshi et al., 2016; Varol et al., 2017).

Several literary bot accounts have posted that they were moving to Mastodon, an open-source self-hosting micro-blogging platform, following Musk's acquisition and prior to the API changes. Between users who saw their account being suspended as they were violating the updated Twitter rules, and those who cannot continue to operate using the free API, Twitter appears to be a hostile environment for literary bots, as well as many other niches of users. The exodus towards Mastodon is supported by the development of new tools. The farewell message on the website of *Cheap Bot, Done Quick!* mentions *Cheap Bots, Toot Sweet!* a new service operating using the same syntax on Mastodon, as a solution to move bots to the alternative micro-blogging platform. Alternatively, actions to "rescue" Twitter literary bots are also undertaken. For instance Flores produced standalone web versions of some of his Twitter bots, previously powered by *Cheap Bot, Done Quick!* (Flores, 2023). The current inhospitality of "X" suggests that the era of Twitter bots as we once knew them has ended. Nonetheless, their legacy lives on on other platforms and through archival projects.

3 Existing typologies for Twitter bots

Literature on literary Twitter bots is scarce. Flores, notably through his "I love e-poetry" blog is the leading scholarly voice on this topic.⁶ In a book chapter dedicated to this form, Flores (2021a) proposed the list of nine bot subgenres quoted below:

- Chatterbots are interactive characters.
- Open/Green bots search through endless data sources and act upon the results to produce their output.
- Closed bots work their way through a finite corpus.
- Ebooks bots publish random samples from a static or dynamic corpus.
- Markov bots generate texts based on a probabilistic analysis of a textual corpus of static or streaming data
- Template bots generate texts by filling in blanks in phrases or sentences.
- Mashup bots combine work from different sources.
- Emoji bots assemble pictorial art and narratives from emoji (small images that act as ideas and are deployed as textual objects).
- Pseudo bots involve partial generation with human curation, or human bot-like performances.

In this typology, Flores focuses on the primary feature of the bot, while acknowledging that the list is not exhaustive and that many of the bots have characteristics of several subgenres (Flores, 2021a). Our own typology builds on Flores' and attempts to isolate different features of the bots (the type of source, the treatment of the source, the interactivity, the variation, etc.) in order to document the full spectrum of bots and avoid reducing them to one of their features.

⁶ URL: <http://iloveeepoetry.org/>

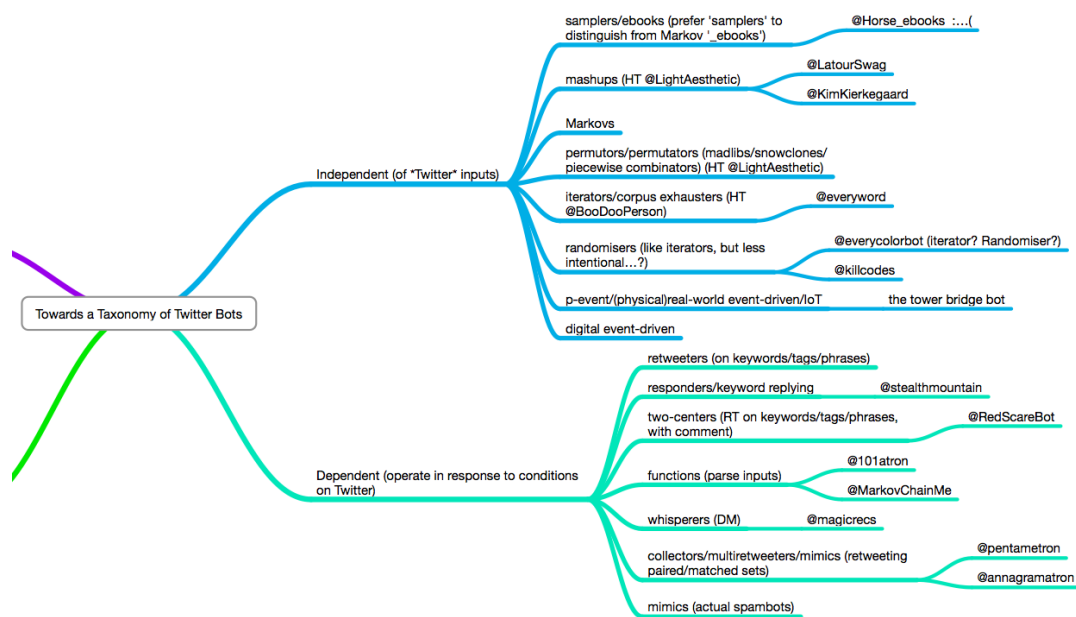


Figure 1: Twitter bot typology by Tully Hansen.

Hansen (2013) proposed another typology which divides bots into two main classes: those which are independent of Twitter inputs, and those which operate in response to Twitter inputs (see fig. 1).⁷ Within those two classes, Hansen proposes subclasses which describe the behaviour of the bots. Although this division is highly relevant in the case of Twitter bots, the binary classification suggests that those two classes exhibit a different set of behaviours. However, there is a large overlap. In addition, bots which rely on another open source accessed through an API are not taken into consideration. This typology illustrates that privileging one of the bot features (here, whether it reacts to Twitter data) over others (e.g. the ways in which it parses input) can be useful to make a specific reading of a set of bots, but also reduces the efficiency and versatility of the typology. To allow for a greater granularity in comparison to Flores' and and greater versatility in comparison to Hansen's, our typology combines multiple categories.

4 The LABEL literary bot typology

By nature, bots can be dissected into a series of formalised steps revolving around three axes: the input (Twitter, other online sources, static texts/data or a combination of the three), the process (ranging from selecting one line from a file at random to a complex algorithm taking several sources as input), and the output (the tweet). The eight fields/categories of this typology describe elements from these three axes. "Source number" and "source type" describe the input ; "procedure", "script number", and "generation axis" describe the process ; and "interactivity", "lexical constancy" and "syntactic equivalence" describe the output. Although each category has a limited number of classes (in fact, six of them are binary classes), they can be combined in many ways. If all combinations were possible—which is not quite the case—we would arrive at 1,024 combinations. Because of this modularity, the typology captures both

⁷ This figure was posted by Hansen on imgur. URL: <https://imgur.com/bKXNQ0V>

the similarities and the differences between bots. Within a collection of bots, we are able to quantify the rarest combination of classes, the most popular, and everything in between, as well as identify the area where any two bots—or groups of bots belonging to the same set of classes—converge or diverge.

Procedure	Source number	Source type	Script number	Interactivity	Lexical constancy	Syntactic equivalence	Generation axis
Quote	Single	Open	Single	Yes	Yes	Yes	Syntactic
Transformation	Multiple	Closed	Multiple	No	No	No	Lexical
Template							Semantic
Algorithm							Metric
							Intertextual
							Statistical
							Orthographic
							Graphic

Table 1: The LABEL literary bot typology

Although only one label per category would usually apply to each bot, potential exceptions are highlighted when relevant. As mentioned in the introduction, this typology focuses on the formal elements of bots. Therefore, using this typology encourages the annotator to trace back the processes which allow the bot to function. In essence, without having access to the code, the annotator infers parts of its structure.

4.1 Procedure

- Quote: The bot quotes existing material *verbatim*.
- Transformation: The bot quotes existing material after operating a transformation.
- Template: The bot generates tweets according to a template. This can be by filling in blanks in an existing phrase or sentence, or generating entire tweets by concatenating a sequence of phrases and/or words according to a pre-established structure. These phrases/words are generally selected at random from a list.
- Algorithm: The bot generates tweets following an algorithmic procedure which is not captured by the three previous procedures. In practice, algorithm bots tend to be more complex, as the generation procedure involves probabilistic or machine learning processes.

4.2 Source number

- Single: The bot relies on one source.
- Multiple: The bot relies on multiple sources.

4.3 Source type

- Open: The source is an open corpus. The most popular open corpus is Twitter itself, but some bots crawl other areas of the Web.
- Closed: The source is finite. This can be a closed corpus (e.g. the complete works of Shakespeare), or a lexical resource (e.g. a list of all adverbs in the English language, or a list of sci-fi film titles).

If the bot relies on both open and closed sources, it can be labelled with both classes. Alternatively, a "mixed" label can be introduced.

4.4 Script number

- Single: All tweets from the bot account are generated according to the same formula, or script.
- Multiple: All tweets from the account are not generated according to the same formula. This is often the case for template bots, for instance, where tweets are sometimes generated according to dozens of different templates. We consider that a bot uses different scripts where the first step in the generation process appears to be choosing between several procedures.

It would be possible for a bot to have different scripts whose characteristics would be different. For instance, we can imagine that a template bot could use one script with a lexical constant and one without (see 4.6). More drastically, a bot could have different scripts which would generate texts according to completely distinct procedures (e.g. a quote and a probabilistic model). Those cases are rare since the interest of the bot is often tied to the single concept it relies on. If needed, we could imagine that scripts could be annotated, instead of bots, or that multiple labels would be used to annotate such accounts.

4.5 Interactivity

- Yes: The bot performs retweets, tweets in answer to other tweets or tags other accounts.
- No: The bot does not directly interact with other tweets or Twitter accounts.

4.6 Lexical constancy

- Yes: All tweets issued by the same script share at least one lexical constant (i.e. they share at least one token).
- No: All tweets issued by the same script do not share a lexical constant.

4.7 Syntactic equivalence

- Yes: All tweets issued by the same script share one syntactic structure. We do not consider quoted elements which do not alter the sentence structure to assess syntactic equivalence. If a script includes film or book titles within the structure of the sentence, for instance, we will consider that the syntactic equivalence is preserved even if the titles have different syntactic structures. We also choose to consider omissions as valid options. For instance, if all tweets have the same syntactic structure, but some omit an adverb, we still consider that they are syntactically equivalent.
- No: All tweets issued by the same script do not share one syntactic structure.

4.8 Generation axis

The generation axis describes the linguistic focus of the text generation procedure.

- **Metric:** The generation revolves around metric. This label is typically used for transformation bots which work their way through a corpus to find tweets or sentences which can be reworked according to a specific meter, such as a haiku.
- **Syntactic:** Tweets are generated following syntactic constraints. For instance, this label is used for template bots which rely on syntactic categories to create syntactically sound sentences.
- **Intertextual:** Intertextuality is the driving force of the bot. One could argue that all literary Twitter bots are intertextual, as they often repeat patterns and structures linked to literary and artistic traditions. However, we use this label for bots which explicitly showcase traces of one or several specific authors or works.
- **Lexical:** The bot acts or relies on lexical elements. For instance, this label would apply to a transformation or quote bots which acts on tweets or sentences which contain a specific word.
- **Semantic:** Semantic relations are at the center of the generation process. This label would describe a template which uses the semantic relations of a resource like WordNet, or lists of semantically related phrases (e.g. film titles, list of colours)
- **Statistical:** The bot uses count-based information to generate text. This label could be used for a bot generating tweets based on the probabilistic analysis of an existing corpus.
- **Graphic:** The visual element of writing influences the text generation procedure. This label is used when tweets create specific shapes or takes into account visual elements such a word length.
- **Orthographic:** The process(es) enacted by the bot rely on spelling.

A bot can work around several linguistic axes. For instance, a bot which generates texts following probabilistic distributions of Shakespeare's plays is both statistical and intertextual. Similarly, a bot which tweets lines from Shakespeare's play if they follow a specific syntactic structure or contain a specific token is intertextual, but also syntactic or lexical, respectively.

5 Case studies

In this section, we illustrate how the typology applies to a few representative examples of literary Twitter bots (see Table 2). Through these examples, both the strengths and limitations of the typology become evident. On the one hand, the combination of the different categories allows us to account for a wide variety of configurations. On the other hand, reducing the complexity of all literary bots to a small number of labels remains illusory, as illustrated by cases which defy this categorisation. In addition, the case studies provide an overview of the main procedural mechanisms at play in literary Twitter bots

User name	Procedure	Source number	Source type	Script number	Interactivity	Lexical constancy	Syntactic equivalence	Generation axis
@mybestdayever	Quote	Single	Open	Single	No	Yes	No	Lexical
@carsonbot	Quote	Single	Closed	Single	No	No	No	Intertextual
@accidental575	Transformation	Single	Open	Single	Yes	Yes	No	Intertextual
@BigramPoetry	Transformation	Single	Open	Single	No	Yes	No	Lexical/Metric
@regrettoegret	Transformation	Single	Open	Single	No	Yes	No	Lexical
@SubstitutionBot	Transformation	Single	Open	Single	No	No	No	Lexical
@gutenboy2bot	Transformation	Single	Closed	Single	No	Yes	No	Lexical
@SpellerBot	Template	Single	Closed	Single	No	Yes	Yes	Orthographic
@is_like_a	Template	Single	Closed	Single	No	Yes	Yes	Lexical/Semantic
@IsItArtBot	Template	Single	Closed	Multiple	No	Yes	Yes	Syntactic
@WhitmanFML	Template	Multiple	Open/Closed	Single	No	No	No	Intertextual
@SnowballPoetry	Algorithm	Single	Closed	Single	No	No	No	(Ortho)graphic
@a2b_bot	Algorithm	Single	Closed	Single	No	No	Yes	Semantic
@AutoImagist	Algorithm	Single	Closed	Single	No	No	No	Graphic

Table 2: Labelled case studies

5.1 Quote bots

@mybestdayever crawls Twitter for tweets including the phrase "best day ever" and copies them in a new tweet. It has a single open source (Twitter) and its generation axis is lexical. @carsonbot tweets a quote from the Canadian poet and essayist Anne Carson every two hours. Its source is a single closed corpus (Anne Carson's work) and its generation axis is intertextual. Although they are characterised by the same overarching procedure, since they both quote existing text *verbatim*, these two quote bots adopt radically different creative processes. One relies on an open corpus and the other on a closed one. One relies on a lexical condition, which leads to the presence of a lexical constant between all tweets, and the other tweets sentences at random, as long as they belong to the chosen corpus. This modular typology allows for these nuances to be captured while acknowledging the commonalities between the two bots.

Quote bots, by definition, have a single source per tweet. If two *verbatim* quotes are combined, we are facing a type of template, since the modality of the combination has to be defined. We could imagine a bot which quotes randomly from one source or from another. Whether we would consider them to be multiple sources or multiple scripts is subject to debate, since we do not have access to the source code and files. This is a recurrent issue in this exercise. Often, the same outcome could have been achieved with different procedures. On the one hand, in digital literature, and especially in literary text generation, where the materiality of the creation process is central, the implications of these nuances can be key to read the creative impulse of the bot. On the other hand, the relative opacity of some Twitter bots is also part of their creative identity. When the creation process is not fully transparent, the ambiguity of the account forces the reader to pause and consider the purpose and effect of the generated text, thereby emphasising a defamiliarisation effect at play when literary content is published on social media platforms.

5.2 Transformation bots

Most transformation bots crawl open corpora and tweet sentences/phrases if they match a selection criterion, after operating a transformation. A popular selection criterion is metric. @accidental575 retweets tweets which can be transformed into



Figure 2: Examples of quote bots.

haikus, i.e. tweets which contain 17 syllables that can be written with a three-phase format distributed in a 5-7-5 pattern. In the quote tweet (retweet with a comment), the bot adds line breaks to transform the tweet into a haiku, credits the original account and adds the hashtag #accidentalhaiku, which becomes a lexical constant. This bot has a single open source (Twitter) and is interactive since it uses the "quote tweet" function and tags the source account. Several accounts works with very similar principles, like @youinhaiku, @HaikuD2, and @haikuincidence.⁸ Various accounts apply the same principles to other sources, whether closed (e.g. @SimpsonsHaiku, which mines scripts from *The Simpsons*), or open (e.g. @WikiHaikuBot, which mines Wikipedia, @nythaikus, which mines the New York Times and @HaikuNewsBot,⁹ which mines 50 English news sources using the free News API).¹⁰



Figure 3: Examples of transformation bots with a metric generation axis.

@BigramPoetry is an example of a transformation bot which works with another metric, since it turns tweets which contain the #machinelearning hashtag into bigram poems.¹¹ In this case, the generation axis is ambiguous since the selection criteria is lexical but the transformation is metric, according to a loose definition of metric. Apart from this additional generation axis, and the lack of interactivity, it has the same

⁸ URL: <https://github.com/warmlogic/haikuincidence>

⁹ URL: <https://roche.io/2017/07/haikunewsbot>

¹⁰ URL: <https://newsapi.org/>

¹¹ URL: <https://dev.to/tomweinandy/building-a-twitter-bot-in-python-to-write-bigram-poems-1dcj>

combination of labels as @accidental575.

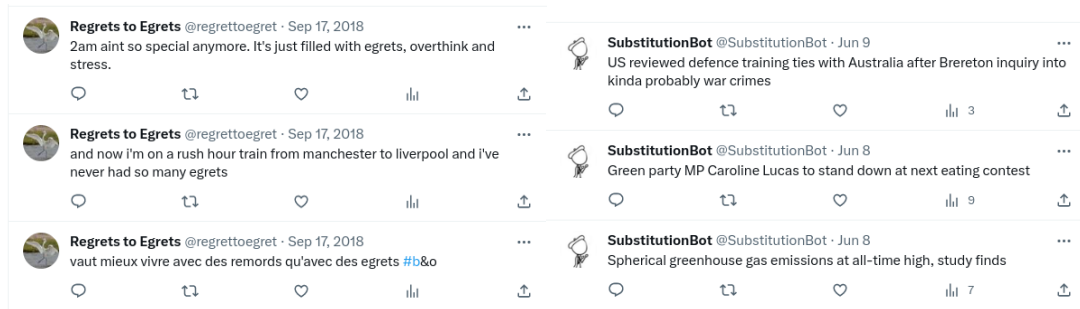


Figure 4: Examples of transformation bots with a lexical generation axis.

Mining tweets for a specific lexical element, which is then swapped by another is a popular process as well. @regrettoegret mines tweets containing the word "regret" and replaces it by "egret". @SubstitutionBot substitutes specific phrases by others in headlines from the Guardian accessed through their API.¹² @gutenboy2bot similarly replaces "boy" with "bot" in classic literature found on project Gutenberg. These three bots are very similar, but the typology allows for their different features to be documented. The generation axis is lexical and they all work with a single source and script. However, the variety of phrases being swapped by @SubstitutionBot is documented by the lack of lexical constancy which is present in the other two bots and the fact that @gutenboy2bot works with a closed corpus instead of crawling an open source through an API is documented in the source type category.

5.3 Template bots

Templates bots are by far the most common literary Twitter bots. Some are very simple and tweet a single phrase with a one-word variable (e.g. @makeamericabot, @dstroyeveryword, @wearable_things). However, most have multiple interrelated variables. This is the case of @SpellerBot and @is_like_a, for instance. The first tweets the phrase "Can't spell X without Y.", where Y is a substring of X, and both X and Y are existing wordforms. The second tweets the phrase "X is like Y since they are both Z", where ZX and ZY are both collocations. These two examples are based on specific lexical resources. Both bots rely on one closed source and one script, and show a lexical constant and syntactic equivalence. The only difference is the generation axis, which is orthographic for the former, and lexical/semantic for the latter. The typology recognises that the formal processes are essentially identical, but that the bots revolve around different linguistic aspects.

For the bots mentioned in the previous paragraph, the combination of lexical and syntactic constancy is a central feature. However, many bots, on the contrary, play with a wide variety of templates so as to "hide" to an extent the bot-like nature of the generation. For instance, @IsItArtBot, uses more than a dozen templates to ask whether an object or a concept is art. The object/concept in question appears to be any possible noun, although we do not know where the vocabulary list comes from. This bot has one closed source (the vocabulary list) and several scripts (i.e. several templates, in this case). There is a lexical constant and equivalent syntax in all tweets generated by the same script, but syntactic equivalence is not maintained across the

¹² URL: <https://github.com/molly/SubstitutionBot/blob/master/substitutionbot.py>



Figure 5: Examples of template bots.

different scripts/templates. Although multiple scripts are most common in template bots, since they multiply the "randomness" characteristic of the procedure, where a simple pattern produces a large variety of different outputs, they can occur in any of the four overarching procedures.



Figure 6: Examples of template bots.

Finally, whether they are based on one or several scripts, many template bots use entire phrases as variables instead of single words. For instance, @WhitmanFML combines lines from Walt Whitman's *Leaves of Grass* with tweets that use the #FML hashtag. The ways in which the two are combined is not immediately evident, but the snippet of Python script in the cover picture of the account and a careful examination of the output suggests that the bot looks for sentences with a comma or conjunctions (such as "and", "but", "or" and "either") around the middle of the sentences/tweets. We can assume that this was only done once for *Leaves of Grass* while the #FML tweets are regularly crawled among recently posted tweets. It picks randomly whether to start the tweet with a Whitman sentence slice or a #FML tweet slice and one sentence from each list. The first sentence is read up to the comma/conjunction (excluded) and the second is appended from the comma/conjunction (included). The bot is based on one open source and one closed source, does not show lexical constancy (since not all tweets include the #FML hashtag) or syntactic equivalence. The generation axis is intertextual, as well as syntactic, since the two variables are united by a conjunction or a comma. This example is different from the other template bots, since the sentences produced by the same script are not syntactically equivalent, as we are not facing a "fill the blank" type procedure. Although the tweets are still produced sequentially via a

template with two variables, these variables are truncated according to a rule before being slotted in. For this reason, it could be argued that this bot is not truly a template bot, but an algorithm bot, since an extra rule is needed after the bot randomly selects the variables from a list.

In Flores' typology, mash-ups are their own category. The decision to abandon this class boils down to the focus of this typology on formal elements. Mash-ups can be achieved by template bots or algorithm bots. In addition, a bot which mixes several sources which do not belong to "recognisable" sources such as Walt Whitman's *Leaves of Grass* and the #FML phenomenon would not be readily considered a mash-up. This is because the label alludes to a political and artistic tradition, in addition to the act of mining different sources. However, intertextual templates with multiple sources and no syntactic equivalence or lexical constant are highly likely to be examples of mash-ups, in contrast to syntactic templates, for instance.

5.4 Algorithm bots

The most popular generation algorithm for literary Twitter bots is the Markov chains algorithm, a count-based predictive model which generally generates each word based on the probability that it follows the previous one. Markov bots have been trained on the Bible (@rewrittenbible, @KJVBlockchain), song lyrics (@MtnGoatsMarkov, @horse_bluegrass), literature (@Fakespearean, @EdgarAllanBot), book titles (@MarkovBooksBot), and tweets from other accounts (@Lin_MarkovBot, @MarkovFollowers), to name a few.



Figure 7: Examples of algorithm bots implementing a version of the Markov chains algorithm.

Some bots combine this procedure with others. @SnowballPoetry generates snowball poems, i.e. poems in which each line is a single word, and each successive word is one letter longer. The two-step program is detailed in the README file of the GitHub repository linked to the account.¹³ First, the input files (from project Gutenberg) are preprocessed: the program looks for any phrases in which the length of each word varies from the previous word by one letter, e.g. "his face", "in the land" and saves each phrase as a separate line in a preprocessed "corpus" text file. Then, words which are not part of a validation lexicon list (Spell Checker Oriented Word Lists) are removed. The bot works by starting with one letter words and randomly traversing a Markov tree that links the second word of one pair to the first of another if they are the same word, until it reaches a dead branch. This bot illustrates the ambiguity of "sources". Are we talking about the raw text files from project Gutenberg, or the preprocessed corpus? Do text files from project Gutenberg count as one or several sources? In this study, we

¹³ URL: <https://github.com/nossidge/snowball>

opted to consider that, in such cases, the text files issued from project Gutenberg are a single source, since they are essentially treated as one text. Preprocessing is part of the bot activity, and, therefore, we do not consider the processed corpus as the source. Nevertheless, we rarely have as much information about the processes, and some bots rely on sources preprocessed by third-parties.

@a2b_bot is a bot created by German digital artist Mario Klingemann. The bio reads "A bot traveling on an associative route through word2vec space. Written by @Quasimondo. #botALLY". The bot uses the word2vec algorithm, which relies on a neural network model to learn word associations from a corpus of texts. This algorithm can identify the semantic nearest neighbours of a given word in a corpus. We can infer that the bot, given two words, finds the shortest path between them in the semantic space created by the word2vec algorithm, taking into account the n nearest neighbours of each word, or setting a maximum semantic distance between each word included in the path. This bot has a single closed source (i.e. the corpus it was trained on). According to our definition, the syntactic equivalence is maintained, although there is no syntax to speak of. The generation axis is semantic.

The bio of @AutoImagist reads "Generating imagist-style poetry by combining random picture descriptions, powered by Microsoft Azure. A bot by @zachwhalen, #botALLY". The bot uses Microsoft Azure's Computer Vision API to describe random photos from Flickr, combines two or more of those descriptions to make similes or juxtapositions, then formats it like a poem "in the style of imagist poets like William Carlos Williams".¹⁴ The bot uses a single open source: the image hosting website Flickr. This source has the particularity of not being textual, ahead of the transformation operated by Microsoft Azure's caption generation algorithm. In addition, this generation algorithm was itself trained on a dataset. The generation axis is primarily graphic, although the text generation itself relates to the semantic axis, to a large extent, since it aims to extract meaning from a picture. This last example highlights that the concept of "source" is ambiguous in more ways than one. Recent machine learning models are often trained on a huge dataset which becomes a source for the text generation procedure. As bots leveraging pre-trained Large Language Models become more common, distinguishing text generation procedures which rely on specialist corpora which were specifically chosen for the task (e.g. @carsonbot and @a2b_bot) from those which rely on sources which are independent from the bot's specific purpose (e.g. @AutoImagist) will become important.

6 The design process

6.1 Data

The corpus from which this typology was developed was created by manually searching for literary bot accounts in English on Twitter. We adopted an inclusive understanding of literature and included all text-based bots with an artistic/creative component. We retrieved 588 bot accounts and 15 million literary tweets in total (which included non-bot accounts as well). This corpus does not include bots whose output exclusively contains emojis or visual material. The collection took place in January and February 2023.

The corpus is therefore not exhaustive, nor can we argue that it is representative of the form. Crucially, we could not retrieve the numerous literary bot accounts which

¹⁴ URL: <https://github.com/zachwhalen/AutoImagist>



Figure 8: Examples of algorithm bots relying on Machine Learning algorithms.

were deleted or banned before the collection took place. However, considering the sample size, this corpus provides a significant sample of the main genres and types of literary bots.

6.2 Method

Designing the annotation was an iterative process. We started with a baseline annotation scheme with had two categories only: procedure and interactivity. We worked in batches and calculated the inter-annotator agreement for each class. A low agreement meant that the class was ambiguous, and that criteria needed to be refined or that the typology needed to be revised.

This typology became increasingly granular as we encountered the need to differentiate between several types of templates, since they are highly heterogeneous and constitute the bulk of the corpus. We added this granularity by documenting the type of sources and the extent to which the tweets differed from each other. For instance, we wanted to distinguish templates with one-word variables—which output near identical tweets—from templates where variables are entire phrases—which output tweets that have very little in common, both lexically and syntactically. We attempted to assess the proportion of "variables" vs. "constants" in templates, but it proved difficult. We considered focusing on the number of tokens or word sequences which either remained constant or were "filled in" for each tweet. The two approaches led to drastically different template groupings. Many templates contain repeated words. Should these count for one variable/constant or several? In addition, template bots which work with several templates largely complicated the matter since not all templates acted similarly. The issue was partially solved by separating lexical and syntactic similarity, and making the decision to reduce lexical similarity to a binary class (i.e. asking whether there is at least one lexical constant).

Since we worked with template bots to define these new ways of differentiating bots, we initially used them to separate the template class into several different classes. However, the type and number of sources, the number of scripts, and the lexical and

syntactic constancy are also relevant for the other procedures. Therefore, we made the decision to include this information in separate categories and use the procedure category for the overarching four types of procedures only. The generation axis was added last to account for the fact that some bots enact different formal processes but work around the same linguistic axis, and vice versa.

7 Discussion and conclusions

Although Alan Turing's "imitation game" is arguably the starting point of the history of bots, literary Twitter bots rarely aim at passing for humans. On the contrary, they often draw attention to their own procedurality. This typology describes bots as the sum of various formal elements. From a research perspective, this approach can provide a relatively detailed distant reading of a corpus. It allows researchers to observe correlations between different categories, identify trends in the distributions and combinations of labels and, conversely, identify bots which have an uncommon or unique combination of labels.

In addition, this typology is also useful in the context of preservation efforts for digital literature. The history of Twitter reminds us that, in terms of preservation, Internet is not forever. Social media, in particular, is dependent on constantly evolving platforms which cause the content to change, break, or disappear. Moreover, both legal constraints linked to Data Protection and the constant flow of massive quantities of data encourage us to rethink conservation and preservation practices. As argued by Giselle Beiguelman, adequate terminologies and methods are needed to address the cultural output of digital media and the masses of data and memories that "vanish [...] in collective and personal archives on the internet" (Beiguelman, 2014, p. 175). The unpredictability, uniqueness and open-endedness of social media feeds also have implications for the archiving of literary works on social media. A fixed version of the "restitution form" (Bachimont, 2007), i.e. the content as consumed by readers, cannot be captured and preserved. When exposing various approaches to the preservation of digital literature, Bouchardon and Bachimont (2013) argue that preserving a description of the content, instead of the content itself, bound to be incomplete, is the most potent on a theoretical level. A dataset of bot metadata becomes a form of preservation. This typology abstracts the processes dependent on specific files, resources, tools and scripts into non variable elements, e.g. the fixed or open-ended nature of the source material, the transformation/generation processes applied to this material, and the formal variability of the output. In addition to providing tools for analysing literary bots, individually, within a collection or alongside other types of generative digital literature works, it also contributes to a fragmented memory of this form, and its contextualisation within the landscape of digital literary art.

Although bots lend themselves more readily to this type of abstraction than most other types of literature, this reduction of complexity nevertheless favours a specific reading. Recording the content rather than the form, for instance, would lead to a very different clustering of these bots. Is the bot humorous, "feel-good", political/subversive, didactic, poetic, narrative, entertaining, reflexive? Does it revolve around a specific event, a specific theme or a specific fandom?

Even when focusing on formal elements, many other categories could be added to this typology, and, within the existing categories, some classes could be refined. For instance, algorithm bots include a wide range of procedures, from simple Markov chains generators all the way to text generators built from Transformers. This lack of

refinement is due to the comparatively small amount of bots pertaining to this class in our dataset. Since the methodology was built empirically, departing from our data, and not from everything that would be technically possible, the level of granularity offered is tailored to this specific corpus. Nevertheless, as technologies evolve and this typology adapts to other platforms and other datasets, it could be important to specify the nature of the algorithms. Another blind spot of this typology concerns the tools used to develop these bots. For instance it would be interesting to know the bots which used *Cheap Bot, Done Quick!* or the SSBOT tool developed by Zach Wahlen, when this information is available.

In addition, the existing typology also presents some challenges. Namely, when a bot relies on several scripts whose processes do not belong to the same classes, we are forced to adopt a multi-label approach, which implies a loss of precision. Even when bots are limited to one script, they sometimes require multiple labels per category. For instance, many algorithm bots are also templates, since the outcome of the algorithmic process is arranged into an output according to a specific pattern (see @a2b_bo in 5.4). In fact, bots which resist the simplicity of the typology are often the most interesting, since they push the envelope of literary bot formulas. For instance, @AliciaWnderland is a bot which translates the Esperanto translation of *Alice's Adventures in Wonderland* back to English. It is a transformation bot, since it quotes existing material after operating a translation. However, this transformation is much more complex than those seen in any of the other transformation bots present in our corpus. Instead of substituting a word with another, or rearranging the text according to a specific metric, it algorithmically generates a new text from the source text. In this case, whether the transformation or algorithmic procedure has precedence is not obvious. This bot exposes the imprecision of the term "algorithm". Because other transformation bots relied on comparatively simplistic processes, and because other algorithm bots do not take single sentences from existing material as sole input, the ambiguous boundary between these two procedures was not clearly apparent before it was confronted with this specific example. Moreover, the generation axis of this bot is intertextual, but it also revolves around translation, making a case for a new class to be added to this category.

The countless ways in which bots can defy and alter this fragmentary typology obscure the fact that, in spite of this creative diversity, the Twitter environment is highly constrained. The character limit is prescribed, but so is the visual appearance of the tweets (font, font size, font colour, layout). The opportunities for interactions are also limited, since they are circumscribed to the features enabled by the platform, which is not designed for literary processes. The possibilities for user input in the creation process are therefore restricted. The narrow "wobble room" afforded by Twitter facilitated the development of this typology. Since so many elements of the bots are fixed by the platform itself, few variables are left to the creators' discretion. If the platform is so limiting, why has it proved relatively attractive for creators? With these constraints, comes the ease of creating and publishing a literary bot. Thanks to tools like *Cheap Bot, Done Quick!* and the platform itself, a bot could be created with no coding skills, hosted for free for an indeterminate duration, and easily shared with others.

In addition, these constraints also contribute to the continuum between literary bots and the avant-garde movements which inspired many of them, explicitly or implicitly. Dadaist "cut-up" techniques, surrealist "cadavres exquis", Oulipian combinations, and other constraint writing techniques all find illustrations in the bots presented in this

article. This continuity is often acknowledged by the bot creators themselves. @SnowballPoetry explicitly refers to Oulipo, the literary movement that coined snowball poems as part of their constrained writing practice. The profile photo of @IsItArt is Marcel Duchamps's urinal (*Fontaine*). This bot playfully mirrors the dadaist challenge of the accepted definitions of art, and does so by making a type of computationally enabled chance collage, reminiscent of dadaist writing techniques.

The duality of chance and constraint is illustrated by the numerous "accidental haiku" bots mentioned in 5.2, since they work with a constraint which was met by chance. However, this duality of chance and constraint is ubiquitous in most literary bots, since they integrate an element of randomness and work with a set of constraints. We could even argue that it is the case for all tweets, for which the reading, to an extent, is a function of their random location on a timeline and juxtaposition with other tweets. This juxtaposition is the reason why, as a platform which is not created for literature, Twitter is an interesting stage for literary bots. In the context of social media, the cut-up techniques favoured by dadaist poets contribute to a defamiliarisation of such platforms and comments on the "noise" created by the overabundance of published content. The relentless tweeting and the near infinite number of closely related tweets created by literary bots reflect the constant updates of our social media timelines, where the same content appears time and again, with small differences. It is in part because literary Twitter bots reveal their own procedurality that they are able to draw attention to the endless illusion of novelty encouraged by social media platforms.

References

- Wenceslao Arroyo-Machado, Enrique Herrera-Viedma, and Daniel Torres-Salinas. The elon musk paradox: Quantifying the presence and impact of twitter bots on altmetrics with focus in social sciences. April 2023. URL <https://dapp.orvium.io/deposits/644235015db3c5af25159230/view>.
- Bruno Bachimont. *Ingénierie des connaissances et des contenus: le numérique entre ontologies et documents*. Hermès Science, 2007.
- Giselle Beiguelman. *Reinventing Memory is Necessary*, pages 12–33. Petrópolis, Edusp, 2014.
- Serge Bouchardon and Bruno Bachimont. Preservation of digital literature: from stored memory to reinvented memory. *Cibertextualidades N*, 5:184–202, 2013.
- Jean Burgess and Nancy K. Baym. *Twitter: a biography*. New York University Press, New York, 2020. ISBN 978-1-4798-1106-9.
- P. R. Chamberlain. Twitter as a vector for disinformation. *Journal of Information Warfare*, 9(1):11–17, 2010. ISSN 14453312, 14453347.
- Charlotte H. Chang, Nikhil R. Deshmukh, Paul R. Armsworth, and Yuta J. Masuda. Environmental users abandoned twitter after musk takeover. *Trends in Ecology & Evolution*, 38(10):893–895, October 2023. ISSN 0169-5347. doi: 10.1016/j.tree.2023.07.002.
- Nikan Chavoshi, Hossein Hamooni, and Abdullah Mueen. *Identifying Correlated Bots in Twitter*. November 2016. ISBN 978-3-319-47873-9. doi: 10.1007/978-3-319-47874-6_2.

- Heather K. Evans, Victoria Cordova, and Savannah Sipole. Twitter style: An analysis of how house candidates used twitter in their 2012 campaigns. *PS: Political Science and Politics*, 47(2):454–462, 2014. ISSN 10490965, 15375935.
- Leonardo Flores. Artistic and literary bots. *Electronic Literature as Digital Humanities*, page 131, 2021a.
- Leonardo Flores. Third-generation electronic literature. *Electronic Literature as Digital Humanities*, 2:27–41, 2021b.
- Leonardo Flores. Rescued and renewed works: Taco hell menu, tiny protests, juracán borinqueño, protestitas and a bonus, Aug 2023. URL <https://leonardoflores.net/blog/creative-work/rescued-and-renewed-works-taco-hell-menu-tiny-protests-jurakan-borinqueno-protestitas-and-a-bonus/>.
- Tully Hansen. Towards a taxonomy of twitter bots, Nov 2013. URL <https://gist.github.com/tullyhansen/7621632>.
- Jakob Jünger. *A brief history of APIs*, page 17–32. Routledge, London, 1 edition, November 2021. ISBN 978-1-00-302524-5. doi: 10.4324/9781003025245-3. URL <https://www.taylorfrancis.com/books/9781003025245/chapters/10.4324/9781003025245-3>.
- Mark C Marino. *Critical code studies*. MIT Press, 2020.
- Ana Novoa, Susan Canavan, Ivan Jarić, Pavel Pipek, and Petr Pyšek. Musk’s twitter takeover jeopardizes culturomics. *Nature*, 612(7939):211–211, December 2022. ISSN 0028-0836, 1476-4687. doi: 10.1038/d41586-022-04361-5.
- Élika Ortega. From street protest to twitter protest: A review of leonardo flores’s twitter bots. 2020. doi: 10.7916/ARCHIPELAGOS-FVGC-XS31. URL <http://archipelagosjournal.org/issue04/ortega-review.html>.
- Anya Schiffrin. Disinformation and democracy. *Journal of International Affairs*, 71(1): 117–126, 2017. ISSN 0022197X.
- Spandana Singh and Margerite Blase. *Twitter*. Protecting the Vote. 2020. URL <http://www.jstor.org.kuleuven.e-bronnen.be/stable/resrep26363.12>.
- Chris Stokel-Walker. Twitter may have lost more than a million users since elon musk took over. *MIT Technology Review*, March 2022. URL <https://www.technologyreview.com/2022/11/03/1062752/twitter-may-have-lost-more-than-a-million-users-since-elon-musk-took-over/>.
- Vassili van der Mersch. Twitter’s 10 year struggle with developer relations, March 2016. URL <https://nordicapis.com/twitter-10-year-struggle-with-developer-relations/>.
- Onur Varol, Emilio Ferrara, Clayton Davis, Filippo Menczer, and Alessandro Flammini. Online human-bot interactions: Detection, estimation, and characterization. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1):280–289, May 2017. ISSN 2334-0770, 2162-3449. doi: 10.1609/icwsm.v11i1.14871.

Tamara Velasquez. Of peñabots and post truth: The use of bots and trolls in online disinformation campaigns in Mexico during the Peña Nieto administration (2012-2018). 2019.