

Machine Vision Creepypasta

Surveillance Devices in Digital Horror

Team Members

Marianne Gunderson, Wester Coenraads, Marc Tuters, Gaurish Thakkar, Diego Alves and Hana Marčetić

Contents

Team Members	1
Contents	1
Summary of Key Findings	2
1. Introduction	2
2. Initial Data Sets	3
3. Research Questions	4
4. Methods	4
4.1. Timeline	4
4.2. Linguistic analysis	4
4.3. Sentiment analysis	6
4. 4. Qualitative analysis	7
5. Findings	7
5.1. Timeline	7
5.2. Linguistic Analysis	8
5.2.1 Lexical Statistics	8
5.2.2 Semantic Analysis	11
5.3 Sentiment Analysis	13
5.4. Qualitative analysis	14
5.4.1. Naturalisation of technology and its documentative role in horror	15
5.4.2. Technology as an instrument of horror	16
6. Discussion	18
7. Conclusion	19
8. References	20

Summary of Key Findings

Machine vision technologies bleed into the stories we tell, and is represented in popular community-driven fiction online.

While they commonly feature in digital horror, the horror is often not in the technologies themselves, but they bring the horror into our lives.

When technology is the source of horror in the story, it is often related to the autonomous agency and decision making of AI technologies such as smart homes, facial recognition, and AI bots.

1. Introduction

We are increasingly surrounded by technologies that surveil, monitor and/or record our lives in ever more minute detail. Our devices know us, or try to know us, they gather data about us constantly. In the media, as well as in narratives from popular culture, we are frequently reminded how this data is used to profile us, to recognise our patterns and predict our movements and desires. In these public discourses, the proliferation of these machine vision technologies are often framed in terms of fear, worry, or unease, especially in relation to ideas of mass surveillance, totalitarian states, or personal vulnerabilities. This project aims to study how machine vision technologies, broadly understood as “the registration, analysis and representation of visual information by machines and algorithms” (Rettberg 2017, 2) are imagined to be scary, specifically through the genre of digital horror known as “creepypasta”.

Creepypasta is a term used to describe short internet-based horror stories often shared anonymously and copied and pasted from forum to forum. The word is a derivative of “copypasta”, a portmanteau of copy and paste which refers to short snippets of text that were compelling enough to be copied from forum post to forum post, often as intriguing or absurd non sequiturs. Its origins can be traced back to 4chan, the site of origin for countless internet memes as well as transgressive and destructive behaviours. From around 2007 these often authorless snippets and anecdotes designed to terrify its readers have been one of the staples on the site. Since then creepypasta has spilled over into the wider internet, and there are now several websites, youtube channels, and podcasts dedicated to creating, collecting and retelling these stories. Creepypasta has been an object of growing academic interest in recent years, and has been theorized as digital urban legends (Henriksen 2018), contemporary folklore (Tolbert 2015), as an example “the digital gothic” (Balanzategui 2019), or as a fourth generation of digital fiction (Ondrak 2018). The relationship between creepypasta and technology has been of particular interest to several of these studies. Both Henriksen (2014) and Ondrak (2018) draw lines from contemporary creepypasta to the optical illusions produced by eighteenth-century phantasmagoria. Cooley and Milligan (2018) show how many creepypastas explore the “nightmares of technological nostalgia”, reanimating technologies of the past and imbuing them with unsettling agency.

Where previous approaches have focused on the unsettling potentialities of past technologies in creepypasta, this project seeks to look at their representation of current technologies. A significant proportion of these stories feature cameras, tracking devices or surveillance technology such as web cameras, facial recognition apps, home security systems, AI assistants, and baby monitors.

Previous studies on creepypasta have also mainly been on stories which have received a certain notoriety and have been archived on separate sites, while this project seeks to study the stories as they emerge on the forums dedicated to writing and sharing them.

2. Initial Data Sets

The NoSleep subreddit is one of the most prolific and productive creepypasta/digital horror communities. Using 4CAT, we scraped the subreddit for stories mentioning a list of machine vision technologies that we compiled as a group, consisting of 65 terms such as “drone”, “AI”, “GPS”, several variations and types of cameras, among others. We then filtered out the comments as well as any duplicates, and this resulted in a dataset of 23.303 posts and 53.780.347 words.

Based on the list of technologies in combination with a set of relevant terms we created a metric to measure the relative relevance of the stories in the set, with stories prominently featuring machine vision technologies ranked as more relevant. Story relevance for each text was measured with the following formula:

$$relevance = 10 \frac{n_c^2}{|C|} + \frac{n_e^2}{|C|}$$

Where n_c is the number of occurrences core terms in the text, n_e is the number of occurrences of edge terms in the text, and $|C|$ is the total length of the text (in characters). The core terms (e.g. “drone”, “nanny cam”) were used for scraping and very relevant, while the edge terms (e.g. “application”, “code”) were less directly relevant but frequently occur near directly relevant terms.

For our qualitative analysis we used the relevance metric to put together a dataset of the top 500 most relevant stories from NoSleep. From this dataset we excluded any duplicates, any stories that had since been deleted from the forums, and all multi-part stories, for ease of analysis, which left us with a final set of 218 stories for our qualitative analysis. The dataset included the anonymized code for the author, URL of the story, the text itself and measures of text and title relevance, as well as the number of upvotes the stories had received at the date of sampling.

As creepypasta as a genre originates from 4chan, we also attempted to put together a dataset of relevant stories from 4chan’s paranormal imageboard, /x/. The scraped data contained mostly non-creepypasta texts, e.g. discussions about conspiracy theories. Even after extensive filtering in an attempt to obtain a set of creepypasta stories from this data, the dataset consisted primarily of non-creepypasta texts. In the interest of time and because of the great effort required to produce a usable dataset, we decided to exclude the /x/ data from our analyses.

3. Research Questions

The project was guided by the following questions:

RQ1: How do machine vision technologies figure in creepypasta?

RQ2: How are machine vision devices made to be scary in those stories?

RQ3: What do these stories say about our fears and anxieties related to machine vision and surveillance technologies?

4. Methods

The project made use of a mixed methods approach, combining quantitative and qualitative analysis. Firstly, to explore the distribution and frequency of machine vision devices in creepypasta over time, we produced a timeline analysis using python and Pushshift. Secondly, we did a linguistic analysis of the ways in which the machine vision devices figured in the stories. Thirdly, we used sentiment analysis to investigate the relation between fear and machine vision devices in the data. Finally we did a close reading of a set of stories in order to map common themes and plotpoints and how they relate to machine vision technologies. Using digital methods such natural language processing and sentiment analysis together with a close reading of a subset of the data, we were able to both broaden and deepen our analysis, exploring different aspects of our research questions.

4.1. Timeline

We performed a timeline analysis to get a sense of the popularity and subject of the stories we collected. For this, we used a Python 3.7.7 script with v1.0.5 for data processing and matplotlib v3.0.2 for plotting.

Using Pushshift, we determined the total number of posts made to the /r/NoSleep subreddit each month since 2012. We then counted the number of stories in our dataset that mention the technologies we are investigating, grouped by month. This allowed us to calculate the fraction of total stories on /r/NoSleep that were about the technologies. We also split this by technology grouping, allowing us to get an idea of the prevalence of each of the technology groups on the subreddit. We then calculated the 5-month moving average of the data to better show the trends in the data.

4.2. Linguistic analysis

The main focus of the linguistic analysis that has been conducted is RQ1 and RQ2. Our aim was to provide some linguistic information (frequency, specificity throughout time and semantics of associated verbs) concerning the selected devices that represent machine vision technologies in the texts composing the corpus.

The data that has been used in this step is the full No_Sleep corpus, from which we have extracted only the body of each post.

Corpus characteristics:

- 23.303 texts.
- 53.780.347 tokens.

The following tools were used:

- TXM v.0.8.1 for lexical statistics and specificities (Heiden, 2010).
- UDPipe v.1.2.0 using GUM model (v.2.5) for automatic syntactic analysis (Straka et al. 2016).
- Python scripts (Python 3.6.9) for UDPipe data processing.
- Gephi 0.9.2 for visualization of semantic relations (Bastian et al. 2009).

The whole process chain can be schematized as following:

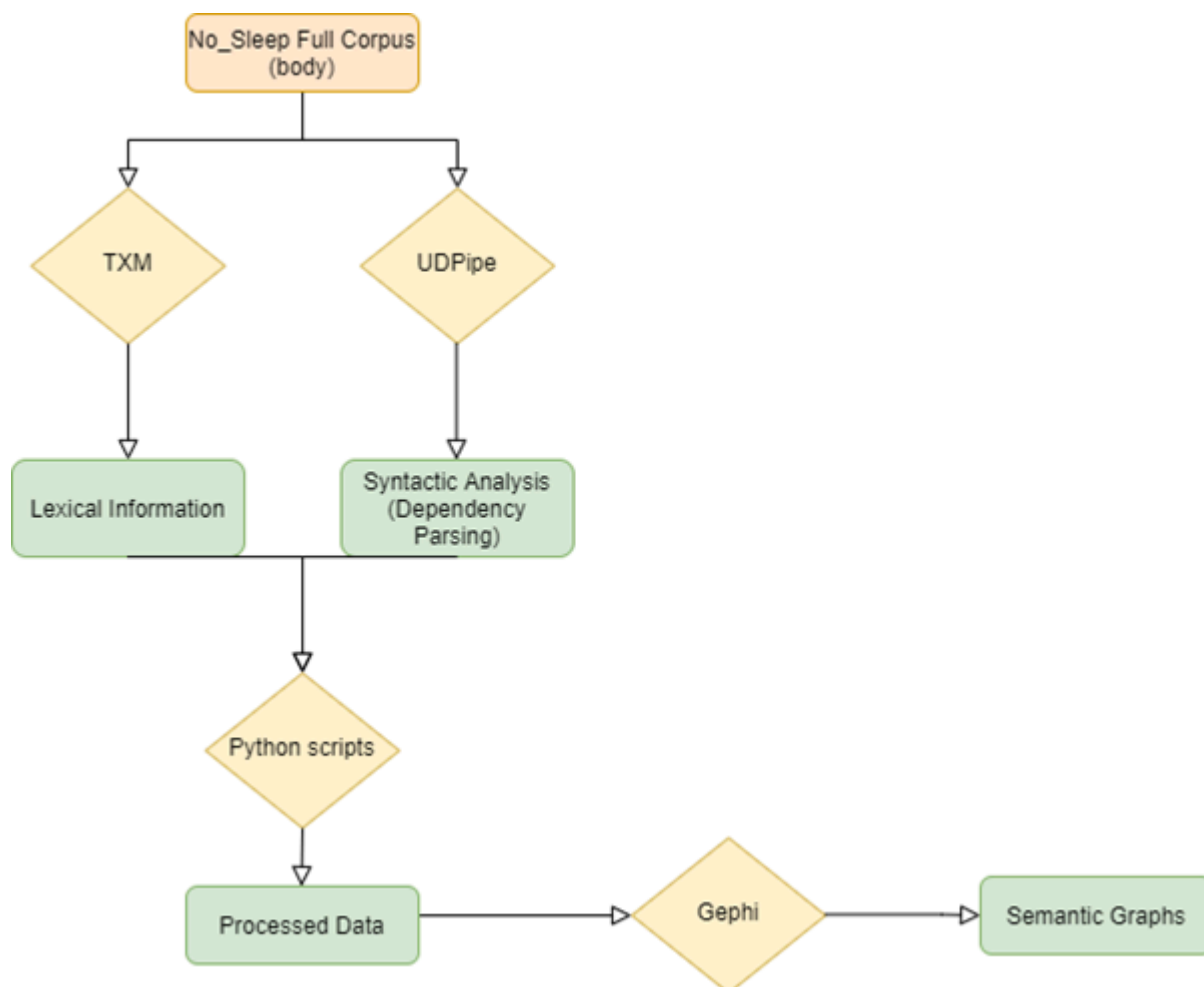


Figure 1. Processing Chain Scheme for Linguistic Analysis of No_Sleep Full corpus.

For the specificity analysis, the idea was to check, for each device, if the term is specific for a certain period of time. Therefore, we have divided the corpus in four sub-corpora:

- Part 1: Texts from 2012 to 2014 - 4. 4.337.188 tokens.
- Part 2: Texts from 2014 to 2016 - 9.567.138 tokens.
- Part 3: Texts from 2016 to 2018 - 12.410.841 tokens.
- Part 4: Texts from 2018 to 2020 - 21.593.367 tokens.
- Part 5: Texts from 2020 - 5.870.063 tokens.

The specificity of a lexicon unit can be defined as the Maximum likelihood/probability of occurrence in a part (TXM Team 2013). Positive values mean that the term is more probable to happen in that specific sub-corpus.

Concerning semantic analysis, we have associated each device characterized as subject or object to its governor (syntactic dependence) to verify if technologies differ in terms of semantic association (particularly with verbs).

For all linguistic analysis, we have used the following list of devices:

list_devices = ["machine vision", "image generation", "deepfake", "facial recognition", "motion recognition", "motion tracking", "motion detection", "detect motion", "AI", "artificial intelligence", "neural network", "deep learning", "algorithm", "biometric", "biometrics", "fitbit", "X-ray", "MRI", "brain scan", "body scan", "body scanner", "VR", "virtual reality", "AR", "augmented reality", "hologram", "siri", "alexa", "google home", "smart house", "smart home", "surveillance", "surveillance camera", "CCTV", "security camera", "GPS", "google maps", "google street view", "streetview", "satellite", "drone", "infrared", "radar", "nightvision", "night vision", "photography", "camera", "camera phone", "cameraphone", "phone camera", "video recorder", "video camera", "nanny cam", "nanny camera", "baby monitor", "webcam", "web camera", "video call", "video chat", "skype", "zoom", "livestream"].

Certainly, this list is not exhaustive, other technologies may be present in the texts as well as typos of the words above that were not considered here. Tokens were considered in upper and lower cases and space has been replaced by underline to guarantee that each multi-token device is considered as one linguistic unit.

4.3. Sentiment analysis

The aim of performing sentiment analysis is to find the relation of emotion fear with the machine vision technologies placed in the stories.

We performed emotion analysis using the tool called [EmoNet](#) on 218 most relevant creepypasta stories. These stories were analysed quantitatively later on. Every sentence from every story was tagged with one of the classes from a list of emotions (joy, anger, anticipation, disgust, fear, sadness, surprise, trust). Since we were more interested in the aspect related to fear, we filtered out all the sentences from the stories that were tagged with fear. We run two analyses on this filtered data. The first question was what is the distribution of fear among the technology-related sentences vs non-technology related sentences. Second, which technology terms are highly associated with fear class among all the fear-tagged sentences.

4. 4. Qualitative analysis

The dataset used for qualitative analysis was a derivative of the top 500 most relevant Creepypasta stories. It was manually scanned for deleted posts and stories that are a part of a series. The remaining 218 stories were the basis for a close reading performed by two members of our team. As many of the stories were of considerable length, we realised that we would not be able to analyse the full set within the constraints of this project, and we decided that an even 100 would serve as a solid basis for a qualitative analysis. In order to achieve a more diverse and relevant sample, we made sure to include stories with a wide range of technologies in our analysis, as well as deliberately including both stories with high and low upvote scores from the forums.

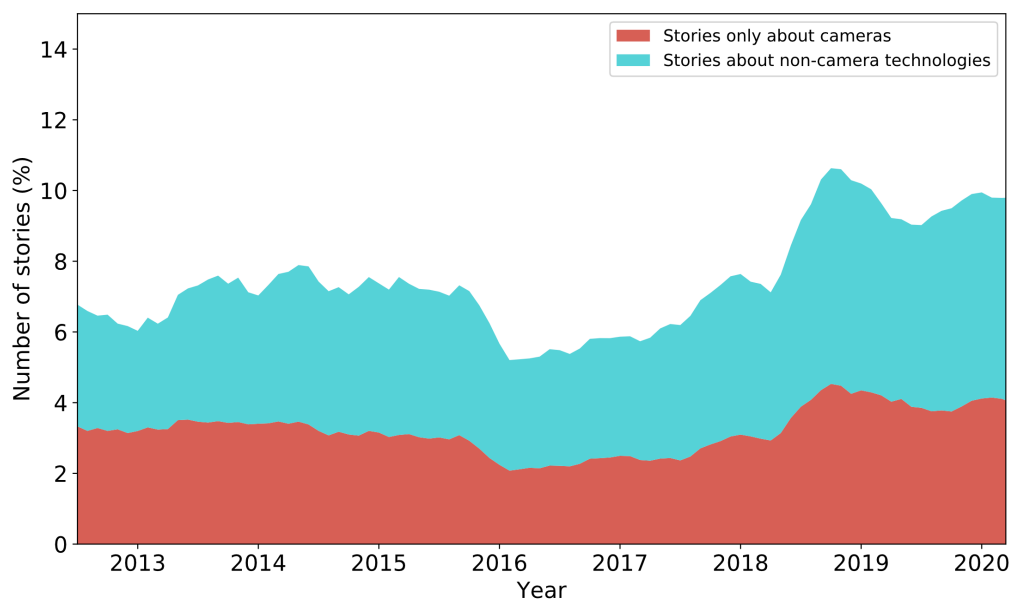
As the project is asking about the relationship between machine vision technologies and horror/fear, our coding system focused on noting which technologies were prominently featured, as well as denoting what the source of the fear or horror was in each story. The coding categories for technologies were taken from the list of technologies used to create the dataset, while the thematic coding categories for source of horror was created through a collaborative grounded analysis, in which now codes were included as common trends were identified. We also identified quotes representing the most common themes, and made notes on especially noteworthy stories. Finally we discussed the full set of 100 tagged stories, drawing on our

impressions and reflections from the close readings and underlying patterns and connections that emerged through the tagging process.

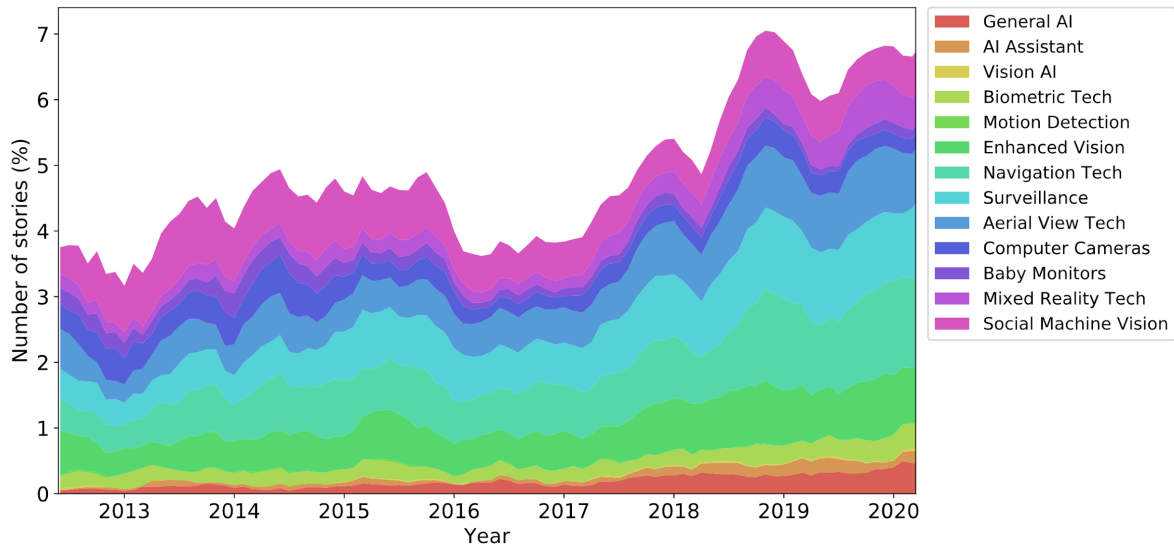
5. Findings

5.1. Timeline

The following figure illustrates the prevalence of the technologies we are investigating on the /r/NoSleep subreddit. The technologies appear in a larger portion of stories over time, reaching about 10% of all stories on /r/NoSleep in 2020. The stories that feature only cameras (and none of the other technologies) have been separated out: stories with cameras constitute a large chunk (between 40% and 60%) of the full set of stories.



The following figure shows a more detailed view of the prevalence of the non-camera technologies over time. Most follow a similar trajectory that is likely shaped by the more general trends on /r/NoSleep. Some technologies increase more than others over the time period: General AI was barely present in 2013 and steadily increased, while Computer Cameras did not increase in prevalence at all.



5.2. Linguistic Analysis

5.2.1 Lexical Statistics

The following table presents the overall statistics of the devices inside the corpus No_Sleep Full:

All Tokens occurrences	53.780.347
Devices occurrences	62.009
Different tokens	214.062
Different devices	62

Table 1. Statistics concerning device's occurrences in the corpus.

It is possible to notice that the devices correspond to 0,12% of the occurrences in the text and 0,03% in the list of different tokens.

The most cited technologies (higher than 1.000 occurrences) are: "camera", "GPS", "skype", "drone", "AI", "webcam", "surveillance", "radar", "satellite", "CCTV", "night vision".

These technologies correspond to 83,4% of all device's occurrences. The token "camera" is the most present one, 55,8% of device's occurrences concern this token. The complete table presenting the occurrences and percentage of each device in the corpus is presented in the link: <https://tinyurl.com/y8k9eax3>.

As mentioned in the previous section, we have also analysed the specificity of each device throughout time (from 2012 to 2020 divided in 5 parts). Using the TXM tool, it is possible to obtain the specificity value of a certain token for each time-period.

The following figure illustrates the results obtained for each device. Positive value means that the token is specific to that period when compared to its occurrences in the whole corpus. In the example below, it is possible to conclude that the probability of the occurrence of the term “webcam” is much higher in Part 1 (2012-2014), decreasing in Part 2 (2014-2016) and becoming negative in Part 4 (2018-2020) and 5 (2020).

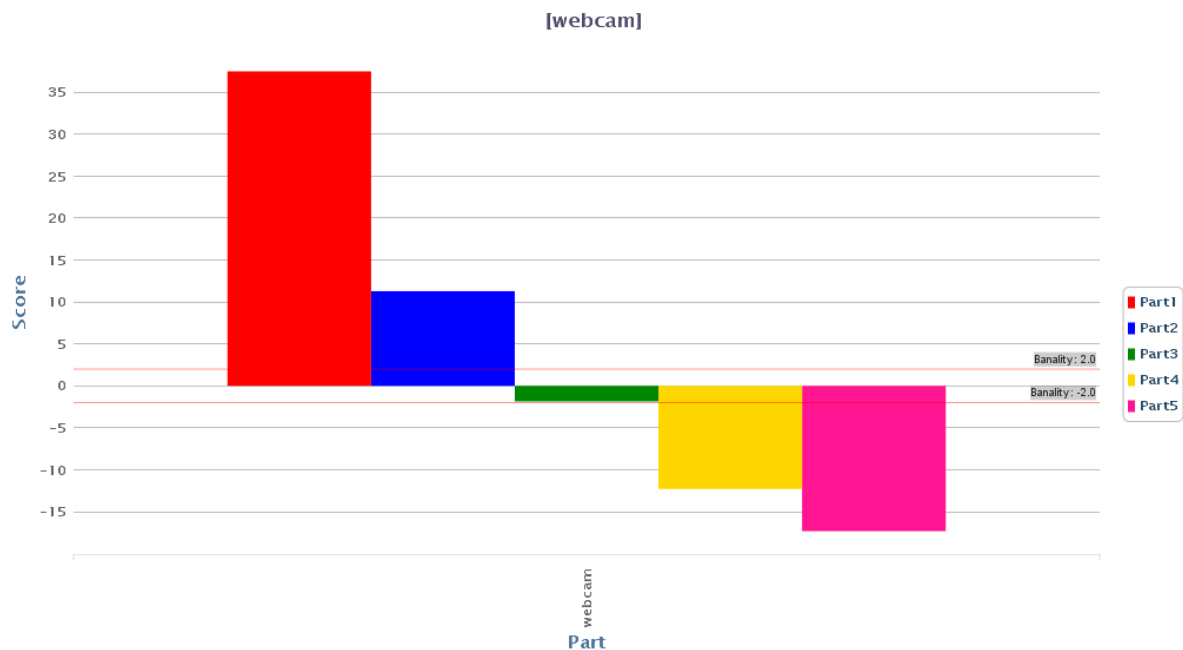


Figure 4: Specificity value throughout time for the term “webcam”.

In the next figure, we present the devices which are specific of each time period:

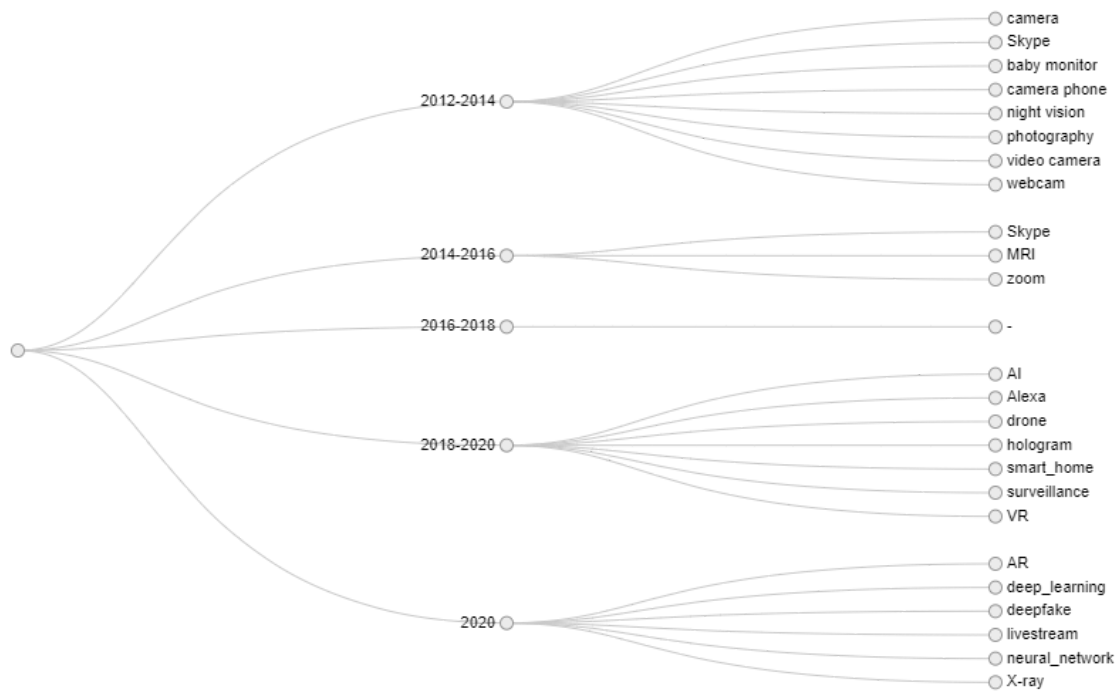


Figure 5: Specific devices for each time period.

It is possible to notice that devices become more specific throughout time, new technologies (for example: "AR", "VR", "neural network") are more probable in more recent texts. It is interesting to see that between 2016 and 2018, there is no device that is more specific. Concerning the listed technologies, this period of time presents a distribution of devices comparable to the whole corpus.

5.2.2 Semantic Analysis

The first step of the semantic analysis is the identification of the syntactic role of the devices in the sentences present in the full No_Sleep corpus. For that, we have used UDpipe tool (Straka et al. 2016) with a model trained with annotated texts coming from different sources (fictional, blogs, etc.).

The following table present the most relevant syntactic roles of the devices:

Syntactic Role	Occurrences	%
obl	14063	24,39333
obj	13691	23,74807
compound	8010	13,89395
nsubj	7845	13,60774
nmod	5405	9,375379
conj	1879	3,259267
root	1505	2,610536

amod	1121	1,944459
nsubj:pass	865	1,500408

Table 2: Syntactic roles of the devices in the No_Sleep full corpus.

The syntactic role labels follow the Universal Dependencies Framework (available at: <https://universaldependencies.org/>). The 9 different labels above correspond to 92,8% of the device's occurrences.

Oblique nominal (obl), which correspond to nominal dependents of verbs that are not subject nor direct object, and Object (obj), direct object, are the most frequent syntactic roles of the devices (almost 50%).

Our focus for the semantic analysis is the verbs associated to the devices, therefore, we have selected on the list above the syntactic relations for which the governor of the syntagme is usually a verb: obl, obj, nsubj (nominal subject) and nsubj:pass (passive nominal subject).

However, to be more precise in the semantic analysis, we have divided these four labels in two groups:

- Subject: nsubj and nsubj:pass.
- Object: obl and obj.

In the first group (Subject), the device is the main actor in the sentence while in the second one (Object), the actor is not the device itself, which is just the complement.

With the syntactic information (label and governor), we have used Gephi software (Bastian et al. 2009) to plot the semantic relations for both groups.

The full graphs are present in the links below:

- Subject: <https://tinyurl.com/ycl5dtfy>
 - Devices: gray
 - Verbs linked to nsubj: blue
 - Verbs linked to nsubj:pass: pink
- Object: <https://tinyurl.com/y8mjzbuk>
 - Devices: gray
 - Verbs linked to obl: green
 - Verbs linked to nsubj:pass: purple

First, we can analyse the position of the devices in the graphs. The size of the circle corresponds to the frequency and its position is related to the semantic classification. Devices which are closer in the graphs are more semantically related (concerning verbs).

Devices such as "camera", "GPS", "drone", and others are more central because of their frequency, therefore, they are more semantically linked to the other devices. In the Subject graph, we can identify three main semantic groups:

- Group 1: "AR", "hologram", "photography", "livestream", "AI".
- Group 2: "satellite", "X-Ray", "GPS", "MRI", "GPS", "CCTV", "radar"
- Group 3: other devices (excluding "camera").

The token "camera" cannot be associated to a group as it is central.

In the Object graph, devices are distributed more homogeneously around the central "camera" token. Nevertheless, it is possible to clusterize the devices in two groups:

- Group 1: "nanny cam", "phone camera", "web camera", "security monitor", "fitbit", "body cam", "surveillance cam", "night camera", "body camera", "night call", "virtual phone", "security scan", "machine intelligence", "motion detection", "hologram", "deep scan", "satellite", "hologram", "CCTV".
- Group 2: other devices (excluding "camera").

The Subject graph allows us also to identify some interesting semantic aspects of certain devices with specific verbal semantic associations:

- "AI" is associated with terms such as "control", "danger", "trick", which can contribute to the horror in the stories.
- "X-Ray" is associated with "revulse" and "crumple" which are usually negative terms.
- Verbs such as "think" and "know" which are usually associated with humans or animals, are associated with Group 3 devices. This personification of the devices can also be a source of horror depending on the context.

5.3 Sentiment Analysis

The following donut chart captures the distribution of fear among the tech vs non-tech story sentences. We see the Fear-Technology count being equal to 1327 and Fear-Not-technology was 4064. Some examples from the distribution are :

1. "The man looks down at her still frightened and his eyes grow even wider when he sees what she is presenting him with." *Non-tech fear*
2. "As I came close enough to the two drones, I saw it turn to me." - *Tech fear*



Figure 6: Fear in technology vs non-technology sentences.

The following table shows the list of most frequent technology terms present in the lines which are tagged with fear as the emotion. The column count captures the normalized value of the terms with respect to the total fearful sentences. The term **camera** occurs with the highest frequency making it the most fearful machine vision technology whereas the least feared is **satellite**.

Terms	Count	Terms	Count	Terms	Count
camera	0.0731	livestream	0.0015	security_camera	0.0005
drone	0.0056	baby_monitor	0.0009	CCTV	0.0003
GPS	0.0056	photography	0.0009	surveillance	0.0003

AI	0.0033	algorithm	0.0007	virtual_reality	0.0003
VR	0.0024	infrared	0.0007	brain_scan	0.0001
webcam	0.0022	video_camera	0.0007	facial_recognition	0.0001
night_vision	0.0018	zoom	0.0007	nanny_cam	0.0001
radar	0.0016	hologram	0.0005	satellite	0.0001

Table 7: Distribution of technology terms and their counts.

5.4. Qualitative analysis

After a close reading and coding of 100 stories, the most common sources of horror identified were *ghosts or monsters* (52 stories), *technology* (33 stories, of which two were coded as surveillance), *violence, murder, death or suicide* (29 stories), *uncertainty or wrongness* (22 stories), and *surveillance, stalkers or being followed* (17 stories). Technology would be tagged as the source of horror in the stories in which the technology was portrayed as the object of the narrator's fear or dread, or when the devices themselves were represented as horrific in some way.

We allowed for multiple tags for each story, and the tags that most frequently co-occurred with "technology" as source of horror were *violence, murder, death or suicide* (14 stories), followed by *ghosts and monsters* (9 stories).

Predictably, common cameras were the most commonly featured technology in the stories included in the qualitative analysis, but the set also included 10 stories about GPS devices, 15 stories about AIs (including smart homes, facial recognition, and AI assistants), 7 stories about VR, and several stories each about specialized cameras such as webcams, baby monitors, or security cameras. Within the subset where we identified the technology as the source of horror in the story, the AI technologies were significantly over-represented, with 13 out of 33 stories.

One of the main findings of close reading the material is the naturalisation of technology in creepypasta stories. While many of the stories drew their horror from well-known tropes such as hauntings, monsters, or violence, machine vision devices often assume the role of facilitator for the story, serve as a portal for the unnatural, or are used as means of documentation of it. The occurrence of video, phone and other cameras far outweighs any other machines and devices mentioned in the stories. Consequently, stories which feature found recordings of *creepy* events were nearly prominent enough in the analysed sample that they could make their own sub-genre.

One third of the stories in our sample featured technology as the main, or one of the main sources of horror. Significantly, in all of the stories included in our set that featured artificial intelligence or algorithms, the technology itself emerged as the source of horror. Virtual reality technologies were also over-represented as sources of technological horror. In these cases, when technology is the element of horror, it often appeared to have to do with the autonomy and agency that the technology assumes. It's often accompanied with the fear of human loss of autonomy and agency on the other end.

Though cameras far outweigh any other machines and devices in the number of occurrences, newer technologies assume the role of source of horror more often. Artificial intelligence (AI),

virtual reality devices (VR), and sometimes sometimes GPS more often assume autonomous agency as the source of freight. Cameras, baby monitors and similar devices usually take on the role of recording and documenting the horror, or serve as a portal (e.g. monsters control it, ghosts appear or communicate through).

5.4.1. Naturalisation of technology and its documentative role in horror

Technology takes on some roles in Creepypasta stories more often than others. In some, devices reveal the main source of horror, acting as a mediator between the world of paranormal and the world of protagonists:

"But the strangest bit, when I zoomed in on the couple at the wedding, was all the extra people in the frame. (...) But there were so many others up there with them...so desolate, so angry. Hanging around at the feet of the groom was this assortment of about ten people with different ages and appearanc[e], just squatting there, all with the same, ragged grey clothes on. And the worst part...the worst part was the eyes. They were all staring at my camera, see."

[My local paper's photographer is retiring - but not because of old age](#)

In a similar fashion, sometimes technology is revealing the monster otherwise unseen:

"Curled up monstrously to sit in the tight space was a creature, and this one I saw in full daylight (...) Nobody in the car reacted. The thing didn't move, just stared at the camera. Taller girl looked at the monster and said, "Shut up, Ashley,"

[Manufacturer recall](#)

Often, though, technology is a gateway to the world of paranormal. The discovery of a USB will lead to being haunted by a wraith, a monster or a ghost:

"After I watched the video, she started showing up everywhere. She comes to my window at night, watching me sleep. I see her in playgrounds, back alleys, out of the corner of my eye. I went to the movies and she was in the front row, but was gone when the movie let out. She won't leave me alone, and I know what I have to do now."

[The Drowned Village](#)

Stories in which technology completely fades to background and assumes a documentative role are by far the most prevalent in the dataset. This is partially due to the number of stories featuring cameras and recordings. In fact, a big subset of analysed stories featured stories that begin with a discovery of recordings. Many of them follow a transcript like style, some of them featuring dates and timestamps of the events, as "transcribed" from footage:

"The following is a transcript of a video recording recovered from a dead body found in the Ottawa National Forest in Michigan's Upper Peninsula. The body, discovered by campers on April 23, 2017, was unidentifiable; police used dental records to determine the identity of the man as Jonathan Kendall, 30, a pharmacist from Flint, Michigan.

WARNING: What you are about to read contains depictions of violence, self-mutilation, and cannibalism. Investigations into the events detailed below are ongoing."

[Michigan Myths: The Little Girl of Ottawa National Forest](#)

In summary, machine vision in creepypasta is often a gaze directed towards the paranormal, in which different devices act as mediators and portals. However, there are stories that assign the role of horror to the technology itself.

5.4.2. Technology as an instrument of horror

A portion of stories features technology as the element of horror itself. Most often the technology breaks away from its expected performance and takes on autonomous agency which induces the creepy in the story. More often than not, the horror stems from artificial intelligence, virtual reality or devices such as drones operated by AI. On top of technology assuming autonomy, element of fright stems from human autonomy being hindered as a result. A human protagonist will often witness AI operated devices making decisions about human lives. Some of the outcomes include AI killing people or taking control of their lives, but one story has an interesting twist. Instead of the horror coming from AI assuming control, in this instance the AI retreats because humans relied on its help too much:

"I guess I'm writing this because I feel guilty. For a year, we've been relying on an AI to do our searching for us. Even after he began to focus on the worst of the worst, we didn't bother to begin doing our own searching. We even laid off our detective because of it. And now, we're stranded. Over 100 cases have come in since the beginning of the year. We don't know what to do- it's piling up and up, and AI hasn't called us since then." I'm just scared that we're messed up. Scared that after all this time, AI has found the weakness of the humans he'd once assisted. The idea that we needed AI to help us."

[An AI has been finding missing bodies. We relied on it.](#)

AI is often implied, and the main technology that assumes autonomy is the device it operates. A recurring theme is drones, operated by AI:

"Autonomy - know what that is? It is when something can make decisions for itself. autonomy is what gives humans free will. Autonomy is behind ever[y] decision ever made by the human race. As for the drones, well, I'm sure you're putting the pieces together."

[The Mind of a Drone](#)

Beside fear of technology that gets out of control, and guilt of relying on technology too much mentioned in one of the quotes above, feeling helplessness in face of technology created is another source of horror, as depicted in the following story:

"Due to the nature of the drone, we could do nothing but watch the live feed of both the drone's camera and its programming's decision making."

[Drone](#)

Though other genres of stories, such as SF and anti utopia, have long employed technology advances as agents of fright, creepypasta authors of Reddit have mostly chosen sticking to assigning technology the role in the background. However, in some instances, the machines and devices themselves play a role in amplifying the horror. The following is a story in which a smart house is used as a vehicle for illustrating the atmosphere of horror:

"Ping.

"Motion detected –"

I tapped wildly at the screen.

Ethan watched, his eyes widening, as each video feed came up. Every single video was of her.

*"Lock the doors!" I yelled, running across the foyer.
 "Your phone wasn't working the other day – we have to make sure..."
 Panting, I jiggled the doorknob. Locked. "It must just be some sort of malfunction,"
 Ethan said, utterly confused.
 "The backdoor's locked, checking the garage now –"
 Ping.
 I pulled out my phone.
 But this time, the notification wasn't from the security system. It was from the occupancy app. My
 finger trembling, I tapped on the icon.
 The text came up, stark white.
 ONE PERSON HAS BEEN ADDED TO YOUR HOME."
[Yesterday my 'smart home' locked me out](#)*

The inherent creepiness of technology itself has been discussed explicitly in one story. Here, the author is analysing the creepiness potential that is incorporated in the technology itself:

*"Google probably knows more about my life than I do. The scary part is not only do I know all this, I am okay with it. One day our kids or grandkids are going to look back at our lack of online privacy and be shocked we let any one company have every detail of our personal lives, and we did it voluntarily. I knew all of this when I clicked the download link."
[I downloaded the info Google has stored on me and some of it is unsettling.](#)*

Interestingly, the overall element of horror in this story isn't the technology at all. The author clearly understood the potential of technology for the format of creepypasta, but opted for a route in which technology is again just the tool. The story unravels to reveal that the storyteller has found a website live streaming his bedroom every night, and a community of people that have watched him sleep for a long time.

In summary, cameras and video recording devices often documents or enables the horror in creepypasta stories. Technologies that are used as inducers of horror are often of newer date, such as AI, VR technology or drones. Those stories that feature technology as the source of the horror, very often use the motif of technology assuming agency and acting in ways that are harmful for humans. Virtual reality that blurs the lines between the physical and the virtual world, artificial intelligence (or drones operated by it) whose autonomy jeopardises or diminishes human autonomy are the most common source of horror among those stories that capture the anxieties devices induce. Interesting examples of infestations and technology transcending its limitations, such as an algorithm that jumps to humans and spreads like a virus, learning from its hosts on the way, are represented in our sample to a lesser degree.

6. Discussion

Machine vision technologies, from common cameras to AIs, have been featured in the stories posted to the NoSleep subreddit since the board was started in 2012, most of the time amounting to 6-8% of the total number of stories on the board. Their proportion has increased over time, at its peak amounting to almost 10% of all of the stories on the subreddit. The most frequently mentioned machine vision technology in these stories are variations on consumer cameras, but over time the proportion of stories that mention AI and VR technologies has increased, something which was further confirmed by the lexical statistical analysis. We further found that mentions of machine vision technologies have become more specific over time, paralleling the increased diversification and societal integration of these technologies.

While any conclusions we draw from the sentiment analysis must be tempered by considerations of the limitations of the available tool, as we were not able to find a tool trained on reddit-based datasets, it is still noteworthy that machine vision technologies were featured in one fourth of all the sentences automatically labeled as “fear”. Within these sentences, cameras was predictably the most frequently mentioned technology, followed by drones, GPS devices, AIs, and virtual reality technologies, despite these technologies being much less frequently mentioned in our corpus. This indicates that in creepypasta stories these technologies are imagined as more frightful than for instance security cameras or satellites.

In general, the close reading revealed that the type of fear-based narratives that media often tell about machine vision technologies, such as mass surveillance, totalitarian state power and control, or commercial or criminal exploitation of personal data, while an element in a few stories, did not appear as a strong tendency in creepypasta stories mentioning machine vision technologies. In quite a few stories the horror involved being watched, followed, or stalked, but in these cases the horror often consisted of the experience of being personally stalked by a person or monster with malignant intent, the surveillance was rarely conducted by a state or corporation, and was not related to a larger system.

More often than not, however, technology will assume the role of documenting the horror or enabling the plot through mediating between horror and the storyteller. In stories that feature machine vision technologies, these monsters often enter the narrators’ lives through these technologies. This was especially common with various kinds of consumer cameras, where the camera or screen here takes on the role of a portal that allows the monstrous or paranormal other to enter our lives.

In those stories that feature devices and technology as a source of horror, in stories where the technologies are portrayed as the source of the horror, it was often in relation to technologies such as artificial intelligence, where those technologies start to intervene in people’s lives in nefarious or violent ways. We found themes of autonomous or infectious technology harming human well-being in various ways (physically or mentally) and technology being manipulated against humans by other natural or paranormal agents the most dominant themes. These stories seem to explicitly address the fear of the power these technologies have over us as they gain ever more abilities and sophistication.

7. Conclusion

This project focused on representation of machine technologies in stories posted to NoSleep subreddit and roles they take on in imagining horror. This allows us to investigate how machine vision devices feature in popular imaginations, and how these devices are seen to channel or focus popular fears and anxieties.

While creepypasta stories largely draw on common horror tropes, they frequently feature machine vision devices in a significant portion of their stories. Facial recognition technologies, drones, GPS devices, smart homes and AI bots are already widespread in NoSleep creepypasta. Our everyday lives are saturated with machine vision technologies, and while this may not yet be apparent in commercially published literature, it is clearly reflected in digital community-oriented digital fiction.

While these devices are often not imagined as horrific in themselves, they often function as a portal through which malicious actors, be they mundane or paranormal, can interfere in people’s lives. This may be traced back to an underlying anxiety about the ever-presence of screens and

cameras in our lives, and the near total access these devices could grant, if someone was determined and able to use them against us.

Finally, the machine vision technologies that are most often imagined to be horrifying in and of itself, are the ones related to artificial intelligence, such as algorithms, facial recognition, and smart home devices. In these cases, the fear seems to relate to the perceived or imagined agency and autonomous decision-making of these devices.

References

Balanzategui, Jessica. 2019. "Creepypasta, 'Candle Cove', and the Digital Gothic." *Journal of Visual Culture* 18 (2): 187–208. <https://doi.org/10.1177/1470412919841018>.

Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

Cooley, Kevin, and Caleb Andrew Milligan. 2018. "Haunted Objects, Networked Subjects: The Nightmarish Nostalgia of Creepypasta." *Horror Studies* 9 (2): 193–211. https://doi.org/10.1386/host.9.2.193_1.

Heiden, S. (2010b). The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme. In Ryo Otaguro, Kiyoshi Ishikawa, Hiroshi Umemoto, Kei Yoshimoto, Yasunari Harada (Ed.), 24th Pacific Asia Conference on Language, Information and Computation - PACLIC24 (p. 389-398). Institute for Digital Enhancement of Cognitive Development, Waseda University, Sendai, Japan.

Henriksen, Line. 2014. "'Come, so That I May Chase You Away!' On Ghost Hunts and Posthuman Ethics." *Somatechnics* 4 (1): 39–52. <https://doi.org/10.3366/soma.2014.0111>.

Ondrak, Joe. 2018. "Spectres Des Monstres: Post-Postmodernisms, Hauntology and Creepypasta Narratives as Digital Fiction." *Horror Studies* 9 (2): 161–78. https://doi.org/10.1386/host.9.2.161_1.

Rettberg, Jill Walker. 2017. "Machine Vision in Everyday Life: Playful Interactions with Visual Technologies in Digital Art, Games, Narratives and Social Media." ERC Consolidator Grant 2017 Research proposal. https://www.researchgate.net/publication/321335312_Machine_Vision_in_Everyday_Life_Playful_Interactions_with_Visual_Technologies_in_Digital_Art_Games_Narratives_and_Social_Media.

Straka Milan, Hajič Jan, Straková Jana. UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia, May 2016.

Tolbert, Jeffrey A. 2015. "'Dark and Wicked Things': Slender Man, the Folkloresque, and the Implications of Belief." *Contemporary Legend Series* 3 5: 38–61. <https://doi.org/10.7330/9781607327813>.

TXM Team (2013). TXM Manual. ICAR Laboratory, Lyon University & CNRS, Lyon, France. URL <http://textometrie.ens-lyon.fr/?lang=en>.