

EARLY VISUAL MEDIA LAB

C I C A N T

filmeu\_RIT  
RESEARCH INNOVATION TRANSFORMATION  
EUROPEAN UNIVERSITIES ALLIANCE



FilmEU\_RIT Ref: 101035820, H2020-IBA-SwafS-Support-2-2020

**Dirk Hoyer** is a screenwriter, filmmaker, researcher, and Associate Professor at BFM in Tallinn. He has directed multiple documentaries and authored award-winning screenplays and academic articles, specializing in storytelling, political imaginaries, and film history. His latest book, "Retopia – Creating New Spaces of Possibility," was published by Routledge in 2023.

ORCID: 0000-0003-1949-5288

dhoyer@tlu.ee

**Tobias Frühmorgen** is a researcher, film lecturer and filmmaker. At Lusófona University he teaches Creative Production and project supervision in BA and MA cinema programs. He is part of a variety of EU / Horizon 2020 / Erasmus+ funded research programs like Crescine, C-Accelerate, Cyanotypes and WIRE. In FilmEU University he is supervising international film projects and is part of the pedagogical development. Tobias is a certified Green Consultant, member of the CILECT Committee for Sustainability and completing his PhD on AI and screenwriting. ORCID: 0000-0001-6133-8302  
tobias.fruhmorgen@filmeu.eu

# THE PENNED PARROT IN THE WRITERS' ROOM FOUR STAGE STORY GENERATION IN COLLABORATIVE SCREENWRITING WITH AI

DIRK HOYER

TLU/BFM

TOBIAS FRÜHMORGEN

Lusófona University / FilmEU / Film University Babelsberg

## Abstract

This artistic research article explores the application of Artificial Intelligence (AI) in the process of screenwriting, focusing on the Four Stage Story Generation method in collaborative screenwriting scenarios. The primary objective of this study is to explore and analyse the dynamics between human creativity and AI capabilities in crafting screenplays. By integrating AI into the traditional screenwriting process—which involves stages of writing, feedback, and rewriting—we investigate the transformative impacts on both the process and the outcomes of screenplay development.

Our research employs GPT-4 to co-create a screenplay for a short film. This collaborative approach highlights the potential of AI to act as a creative partner, contributing to the ideation and development phases of the script. The study is structured around a four-stage model: thematic conception, story world conception, character conception, and dramaturgical conception. Each stage is designed to leverage the strengths of AI in generating content that adheres to established screenwriting paradigms while also pushing the boundaries of traditional creative processes.

The findings from this study reveal that while AI can significantly enhance the screenwriting process by generating innovative ideas and content, it also presents challenges such as maintaining narrative coherence and depth in character development. The interaction between human screenwriters and AI showcases a complex interplay of creativity where AI's contributions are curated and refined by human oversight. This research contributes to the understanding of AI's role in creative industries and sets the stage for further explorations into the capabilities and limitations of AI in artistic collaborations.

**Keywords:** *Artistic Research, Collaborative Screenwriting, Human-Machine-Interaction, GPT-4, LLM, Prompting*

## Screenwriting as Artistic Research

Screenwriting as a creative process is in a permanent tension between art and craft. Established norms of dramaturgy, techniques and toolboxes are in productive interaction with the unpredictable world of ideas, with the complexity of human behaviour and with hard-to-map moments of creative inspiration. In collaborative writing, the concept of the writer's room is employed in an attempt to organise and, on occasion, standardise the commercial workflow efficiency in the ideation and creation processes. However, to date, the processes involved in the creation of stories in artistic contexts have yet to be fully integrated into this organisational structure.

Artistic research as a method of inquiry is thus an adequate methodological and processual approach to reflect what de Assis and D'Errico (2019) describe as 'in-betweenness' which has 'its most powerful force in the middle, in the indeterminate zone of productive tension'. As Bolt (2006) has emphasised the specificity of artistic research lies in a 'double articulation between theory and practice, whereby theory emerges from a reflexive practice at the same time as practice is informed by theory.'

In artistic research, the complexity of the creative process of screenwriting is thus enhanced by the double articulation between theory and practice, its position of in-betweenness. De Assis and D'Errico further delineate the position of the artistic researcher by underlining:

It is then not simply a matter for a practitioner to 'double' as an observer

of his or her own practice or artistic production; rather, throughout its development and renegotiations, practice generates discourse, and can in turn be steered, communicated, and reflected by discourse. (2019)

De Assis' and D'Erricos' analysis of the position of the artistic researcher primarily focuses on visual art or art with a material dimension. The specificity of screenwriting is that the discourse is already inherently verbal. The position of the artistic researcher in screenwriting is thus a dynamic process of script development combined with the self-reflexive stance of a narrative meta-analyst.

In the specific context of this research, the complexity is further enhanced by its focus on collaborative screenwriting (which adds an interpersonal dimension) and human-machine interaction (which adds the dimension of generative AI). The aim is not only to co-create a screenplay with AI but also to examine the process of co-creation from the point of view of artistic researchers. Hence, the 'in-betweenness' is not only between practice and theory, between art and scientific analysis but also in the yet undefined zone between human and AI generated co-creation. Through the integration of AI into the traditional screenplay development approach – writing, feedback, rewriting – the transformations in the process and in the outcome can be examined. The ideation and screenplay development process thus resembles the structure of a writers' room. In this specific experiment the focus is on the creation of a short-film screenplay in collaboration with GPT-4. Artistic research with its focus on the process of creation and its explicit positioning in the indeterminate zone

of productive tension provides a conceptual framework for the analysis of the distortions and the 'anomalies' in the process of script development. As scriptwriters and academics, we thus engage in an open-ended dialogue with the machine. In order to create a bigger challenge for the AI and also specifically examine the artistic possibilities, the prompting parameters for the short-film have been set to non-generic versions of screenplay typically associated with European Art House cinema. The final outcome of the process consists of the first three drafts of a short-film screenplay.

### The Hierarchical Story Generation Approach

Mirowski et al. (2022) introduced Dramatron, an interactive co-writing approach that leverages large language models (LLMs) to generate coherent scripts and screenplays through a hierarchical story generation method. The model used is Chinchilla (Hoffmann et Al., 2022), a specially trained LLM. The co-writing process begins with a user-provided logline, a concise summary of the central dramatic conflict, serving as the seed for the entire script generation. From this logline, Dramatron generates the screenplay's title, list of characters, plot outline (a sequence of scene summaries with settings and narrative beats), location descriptions, and dialogue. The user can intervene at any stage of this generation process, soliciting alternative generations, editing and rewriting output text, or continuing text generation.

The hierarchical generation approach is designed to circumvent the limited context window of LLMs and achieve long-term semantic coherence across the entire script. By decomposing the story generation into layers of abstraction (logline,

character descriptions and plot outline, and dialogue), Dramatron ensures that content at each layer is coherent with content in other layers. The authors demonstrated the efficacy of Dramatron as an interactive co-creative system through a user study involving 15 theatre and film industry professionals. Participants engaged in two-hour sessions where they co-wrote theatre screenplays with Dramatron and provided feedback through open-ended interviews. Quantitatively, the authors tracked writer modifications to generated sentences, measuring absolute and relative word edit distance to assess the extent of the writer's contributions and edits to the AI-generated suggestions. They also employed a Jaccard similarity-based metric on words to quantify the similarity between the draft after edits and the original suggestion. The evaluation process revealed several key findings, including the potential of Dramatron to inspire writers, assist in world-building, and generate content. Participants also noted biases and stereotypes in the generated outputs, as well as limitations in common sense, nuance, and character motivations. Structural criticisms focused on the need for a well-crafted logline and inconsistencies between consecutive scenes due to parallel dialogue generation. Overall, the study demonstrates the potential of Dramatron as a co-creative writing tool that can augment and uplift human artists' work while leveraging the capabilities of LLMs.

### Four Stage Story Generation

The hierarchical generation approach with its focus on loglines does not correspond to our daily experience in screenplay development with students or to our practice as screenwriters. Especially in an artistic context, the results

are generic, lack character depth and originality. Hierarchical generation is also too static for writers' room situations. Four Stage Story Generation, our approach to collaborative screenwriting with AI, is based on the following conceptual stages:

- Thematic conception
- Story world conception
- Character conception
- Dramaturgical conception

The first stage of our model, thematic conception, is commonly a later stage in screenplay development. As Howard and Mabley (1993) point out, experienced screenwriters seldom begin with a theme as this method 'leads to clichés, propaganda, and lifeless characters, because all the human issues of the drama have been subordinated to this thesis the author is out to prove.' Will Dunne (2009) underlines the theme as a key element of stories and differentiates between two primary approaches to story development: intellectual and intuitive. In the intuitive approach, the theme comes after the characters and the dramatic incidents have been developed, and the intellectual approach is premise-driven. In the human AI writers' room, we discovered that neither the intuitive nor the intellectual approach (which is close to Dramaton) are the most productive approaches, but a strong thematic focus at the outset of the idea generation. Trottier (2014) describes the theme as 'the moral meaning'. Sharples and Pérez (2022) define an 'insight into the human condition, a moral, or just a final twist to the plot' as a condition for a satisfactory story. In our collaborative screenwriting with AI the thematic design was the starting point for story development.

Story world conception is the second stage. As Robert McKee (1997) emphasises, a 'story's setting is four-dimensional—Period, Duration, Location, Level of Conflict.' Thus, the story's place in time, the story's length through time, the story's place in space and the story's position in the hierarchy of human struggles are the essential components of the story world conception. Russin & Downs (2012) point out that the story world also provides 'important ambiance, perspective, tone and context in which each particular scene takes place.' The story world conception thus transcends a mere reduction on setting and provides an important factor in creating the fundamentals in which the story unfolds.

In the context of collaborative screenwriting with AI, character development and the human condition, frequently the starting point of story development is only the third stage. Character conception in our human-machine writers' room is based on the fundament defined by Aaron Sorkin (2020): 'A character is born from the INTENTION and OBSTACLE—they want something, and something stands in their way of getting it.' In collaboration with AI another aspect is of primordial importance: the differentiation between character and characterisation. Syd Field (2005) highlights the difference when he describes character as 'the deep-seated nature of who people are in terms of values, actions and beliefs'. In contrast, characterisation is expressed in the way people live and in their tastes. Character conception thus means to avoid a superficial focus on mere characterisation. In our specific context we have concentrated on the characters' hopes and fears in order to make them more complex.

The fourth stage, dramaturgical conception, is focusing on the dramatically effective arrangement of the first three stages. The focus goes back to the etymological roots of the word drama, the Ancient Greek 'dran', to act. The emphasis is on building the conflict through action and reaction schemes based on a well-crafted cause and effect structure. As Cowgill (1999) explains, the dramaturgy guides 'the audience through the story utilising a certain amount of surprise and excitement while at the same time maintaining momentum and meaning.'

In addition to the Four Stage Story Generation, our human-machine collaboration diverges from the hierarchical generation approach in another fundamental way: the story development process is not based on loglines as the seed of the entire process but on step-outlines. McKee (1997) calls this 'writing from the inside out' and defines a step-outline as 'the story told in steps'. These short summaries of the main dramatic action are the structural building pieces of our story development process. Chronologically we proceeded from the ideation process via step outline development sessions to the development of scenes. The final result was the draft of the short film screenplay *Frost Bite*. The work process in our human-machine writers' room was thus based on McKee's 'writing from the inside out' step-outlining followed by an analysis of the output and further recalibration. This story development cycle repeated with a permanent adjustment of the prompting strategy.

### **Case Study *Frost Bite*: The Prompting Strategies**

Our artistic research process focused on the development of a short film screenplay. The outcome of the Four Stage Story Conception experiment with GPT-4 was the screenplay *Frost Bite*. The screenplay tells the story of three members of the female Austrian soccer team who are invited to a nuclear ice-breaker cruise to Franz Josef Land by the Russian influencer Aleksei. Provoked by his arrogance, the women agree on an improvised soccer match on the frozen sea. The increasingly intense game is fuelled by an absurd spirit of competitiveness that ultimately leads to self-destruction. The 13-page screenplay was the result of several months of script development and prompting sessions in our human-machine writers' room. *Frost Bite* provided the case study for testing different artistic prompting methods.

Prompting is a methodology for engaging with LLMs. By crafting instructions to produce useful outputs it inherently brings a multitude of constraints that are intrinsically linked to the way these models have been designed. Firstly, it is crucial to recognise that the initiation of the prompting process necessitates a human-initiated input. This is a direct consequence of the absence of intentionality within the LLMs themselves.

Moreover, the autoregressive nature of LLMs, wherein the generation of output proceeds sequentially, with each subsequent word or token being conditioned upon the previously generated ones, dictates that the prompting procedure must be executed in a step-by-step approach. The inherent inability of GPT systems to engage in proactive planning not only contributes

to their potential instability and the manifestation of hallucinatory outputs but also underscores the need for a structured, incremental approach to prompting. The third constraint arises from the inherent deficiency in world knowledge that characterises these language models. The absence of a comprehensive understanding of real-world concepts and their intricate interrelationships imposes limitations on the extent to which LLMs can autonomously generate coherent and contextually relevant content. In response to the fourth constraint, based on the results from Nelson (2023) that 'current language models do not robustly make use of information in long input contexts' and 'Model performance is highest when relevant information occurs at the beginning or end of its input context', we restricted the input prompts to a maximum length of 200 tokens. In light of these constraints, our prompting principles governing this artistic collaboration are as follow: (1) the initiation of the creative process lies within the purview of human participants, who (2) bear the responsibility for curating and filtering the generated outputs; (3) ultimate decision-making authority rests with the human collaborators, while (4) the machine assumes the role of a suggestive entity, proposing ideas and concepts for further refinement and integration into the narrative work. In the light of co-creation, our prompting strategy consisted of (1) prompting (2) output evaluation (3) recalibration. This process was repeated until the human curation considered it satisfactory.

We used the standard GPT-4 from OpenAI in the browser, without further instructions or tweaking of temperature or the GPT Assistants ('GPTs') with the possibility of uploading own training data. We used this despite knowing that other models might be stronger, more trained towards screenwriting

and ethically less conflicted with their source of training data. Nevertheless, GPT is the most accessible model for a non-technical trained screenwriter.

## Prompting the Thematic Conception

The first stage of our prompting strategy was prompting for a moral meaning. Expanding Trottier (2014), we defined moral meaning as the underlying ethical or moral message conveyed through a story, shaping characters, plot, and conflicts and leading to a deeper understanding of human behaviour and society.

Prompt examples: '*Make a list of moral meanings of films you know*'. Typical outputs for this prompt are: '*Eternal Sunshine of the Spotless Mind: The importance of accepting both the joys and pains of relationships.*' or '*Malcolm & Marie: The moral complexities of love, communication, and the fine line between passion and toxicity.*' or '*Don't Look Up: A call to action for society to take scientific warnings seriously and address global crises.*' It should be noted that none of the screenplays were uploaded to GPT-4 from our side.

Further prompting asking for '*moral meanings for films that don't exist yet*' resulted in '*The importance of resilience and adaptability in the face of life's constant changes.*' or '*The importance of empathy towards other living beings and the complexities of interspecies communication.*'

These results are very conceptual, fulfilling the 'insight into the human condition, a moral, or just a final twist to the plot' (Sharples and Pérez, 2022). The question remains as to the

extent to which they are useful for the creation of films. It is important to note that originality in storytelling is rarely accompanied by absolute consistency and spotlessness; rather, it is often accompanied by contradictions in human nature and the world in which we live.

When prompted more specifically on the basic set-up of our story *Frost Bite* ('people playing soccer on the Arctic ice'), the suggested five proposed moral meanings were 'perseverance in the face of absurdity, the absurdity of competitiveness, human connections in isolation, environmental responsibility and redefined notions of success'. In order to avoid more abstract themes (such as 'redefined notions of success' or 'perseverance in the face of absurdity'), to achieve the maximum amount of direct confrontation (the theme 'human connections in isolation' has a less direct conflict potential) and in order to avoid a too much AI bias ('environmental responsibility' and similar themes typically lead to overly didactic themes) we opted for 'the absurdity of competitiveness'. Our soccer match on the polar ice would thus be structured by what the LLM defined as: *'a commentary on the absurdity of competition itself. It's a tongue-in cheek critique of the lengths people will go to win, no matter how ridiculous or meaningless the objective'*. The theme *'encourages viewers to question the motivations and values that drive their own competitive nature'*.

The further development this thematic framework would guide our story-world conception, the character conception (we focused on the motivations and values that drive the competitive nature) and the dramatic conception (we foregrounded the 'dram' – the dramatic action – in order to show the lengths, the characters go to to win an essentially absurd game)

## Prompting the Storyworld Conception

The second step of our strategy is designing the storyworld with its dimension of period, duration, location, ambiance, perspective, tone, context and level of conflict. This is sometimes also referred to as 'setting' (McKee 1997). In order to ensure the prompts were concise and working against GPT's bias towards Silicon-Valley-based near-future scenarios, we initially investigated its comprehension of the characteristics of 'European Art House Films'. This led to the definition of *'complex existential, spiritual, and political themes related to the human condition, favouring atmosphere, ambiguity, and contemplation over clear plot resolution.'* This provided reassurance that the GPT was capable of producing outputs within the desired genre and tone. The thematic conception ('the absurdity of competitiveness') combined with the basic set-up (people playing soccer on polar ice) created a more narrowly defined framework for story-world conception.

Secondly, we requested the inclusion of scenarios that could be described as 'lose-lose' (also known as moral dilemmas). These scenarios were selected as they promised both an internal and external conflict, as well as offering a first description of a character, as these two are deeply entangled. This scenario automatically included all the elements of the storyworld.

Prompt example: *'Develop three complex moral dilemmas for a character in a contemporary European art house film, based on the theme "the absurdity of competitiveness". This character should have a grounded, everyday profession but face situations that challenge deep personal and ethical convictions, impacting'*

*both themselves and their close relationships. Each scenario should present a choice leading to nuanced, unfavourable outcomes, underscoring the character's intricate moral landscape. Ensure that these dilemmas are unique, not previously explored in film, and craft each scenario to evoke strong emotional and ethical tensions, showcasing the character's depth and the stark reality of their decisions.'*

Typically, this resulted in outputs providing the film's title, setting, short character description, and a dilemma (with choice a and b). That included a precise description of the place and time (e.g. '*Onboard a malfunctioning research vessel in the middle of a stormy sea*'), a title (e.g. '*Artic Ice*'), characters (e.g. 'Dr. Katrina Volkova is a renowned climate scientist from Moscow, Russia. She has spent much of her early career in labs and at conferences, advocating for stronger measures against climate change') and a dilemma ('Alexei approaches Katrina, having recognised her not as the scientist but as the celebrated soccer player...')

At that level, all the output maintains a consistent quality, clearly outlining conflicts on both levels (internal and external). There is a potential for dramaturgical tensions, all fitting well within the overarching thematic framework. However, the results appear to be overly polished and predictable, lacking the unexpected, the irritating, the contradictory, and the human elements that would have made them more compelling for a cinematic experience. The output lacked both the immediacy required for the direct confrontation as well as a consistent thematic focus on the absurdity of competitiveness.

## Prompting the Character Conception

The third step was creating characters (protagonist, antagonist, secondary characters) with a backstory, description of their relationship to each other, internal and external conflict, motivation, and unique twist. As pointed out by Riedl (2010), 'Character believability is the perception by the audience that the actions performed by characters do not negatively impact the audience's suspension of disbelief.' The principal challenge was to establish a coherent link between these disparate narrative characteristics. To achieve this, prompting was done in several consecutive steps following selecting, curating, and refining. This shows an example of the earlier stages of story creation:

Prompt 1: '*create 10 characters that are radically different from each other, but they all have a special relationship, dependent on each other.*'

Output for prompt 1: '... 3. Eleanor "Ellie" Brooks – A 65-year-old retired British naval officer. Stern but fair, she has a wealth of experience and often acts as a mentor to the younger members of the group. 4. Takashi Yamamoto – A young Japanese tech entrepreneur in his early 20s, shy but brilliant, especially in robotics and AI. He often collaborates with the group to provide technological solutions...'

Prompt 2: '*make pairs of 2. Describe how they are related and dependent on each other. create a maximum level of conflict. Include an intriguing setting.*'

Output for prompt 2: 'Eleanor 'Ellie' Brooks and Takashi Yamamoto – Setting: Onboard a malfunctioning research vessel in the middle of a stormy sea. Relationship: Ellie's navigation skills are critical to guiding the ship to safety, while Takashi's tech expertise is needed to repair the ship's failing systems. Conflict: Ellie's traditional approach clashes with Takashi's reliance on technology, creating tension as they race against time to prevent a disaster at sea.'

The co-creation of characters with GPT-4 presents a challenge. While the model demonstrates an impressive capacity to generate extensive character backstories with a large number of meticulously described elements, these elements often lack the depth, coherence, and authenticity necessary to satisfy the demands of believability. The generated details, while superficially interesting, tend to manifest as a disparate collection of random attributes that fail to fuse into a cohesive and compelling whole. This can be illustrated by the following example output: *'Kata is fiercely intelligent, resourceful, and has an adventurous spirit. She possesses a quiet strength and is deeply reflective, often lost in her thoughts about the future of the planet...'*

This inherent superficiality of GPT-4's character creations may find utility in formats of low innovation level (Deuerling, 2016, p. 29), where predictable storytelling is tolerated or desired, such as telenovelas or scripted television; however, it falls short of captivating the discerning tastes of a cinema or streamed-content audience that craves originality, believability, depth, and genuine surprise. Attempts to prompt GPT-4 for 'unconventional characters' frequently result in archetypal artists, scientists or other trope-loaded constructs, revealing

the limitations of the model's creative range. A good illustration is the early-stage character description of the protagonist Alexei: *'Alexei is ambitious, driven, and methodical. He has a relentless pursuit of truth and knowledge, often bordering on obsession. (...) Beneath Alexei's scientific curiosity lies a deep-seated need for redemption and self-acceptance. His experiments on competitiveness are less about the subjects he studies and more about his journey to reconcile with his past.'*

The character description is conceptual, without any emotional attachment and does not allow for any audience identification. The relation to the three Austrian soccer players is described as follows:

*'Alexei feels a strange camaraderie with the Austrian women, seeing them not just as subjects but as kindred spirits in their quest for excellence. Despite his provocative nature, he admires their dedication and sees this experiment as a mutual journey of discovery.'* The exaggerated correctness of Alexei's attitude towards the soccer players and the well-intended 'strange camaraderie' levels down the conflict and thereby the dramatic potential of the story.

Moreover, the structural disconnect between a character's backstory and their actions within the narrative is a striking deficiency that undermines the overall coherence and believability of the generated content. In an earlier version, one of the Austrian soccer players was described in the following way: *'Kata embarks on the touristic journey across the Arctic Ocean on a nuclear-powered icebreaker not just for leisure but as a personal pilgrimage. Her grandfather, who recently passed away, had always dreamt of taking her on this journey but never had the*

chance. *The trip is a homage to him, fulfilling a promise she made as a child.*' The potential traumas or formative experiences of a character, as detailed in their backstory, fail to manifest in any meaningful way as influences on their behaviour or decision-making within the story. This lack of coherence appears to stem not from a deliberate artistic or aesthetic choice, but rather from a fundamental limitation of the model's ability to create truly multidimensional characters. Further compounding these issues is the absence of subtext in the generated dialogue, resulting in exchanges that feel flat, unnatural, and devoid of the nuance and depth that define compelling human interaction. A short dialogue example is given below of a scene that describes the first meeting of Alexeij and Kata (and her teammates) on the icebreaker:

ALEXEIJ

(veiling admiration with curiosity)

Champions of the ice, out in the cold? Must be a story there.

KATA

(her tone light, guarded)

*Every story has its backdrop. This is ours, tonight.*

ALEXEIJ

(his smile more measured, persuasive)

*What about adding another chapter? A game on the ice, perhaps.*

*A friendly match.*

*The women exchange glances, their scepticism barely masked.*

ANNA

(dryly)

*Because frozen wastelands scream leisure.*

On the level of character conception, the main weaknesses are thus the lack of depth, coherence, and authenticity. In addition, the absence of dramatic characterisation skills often produces characters that interact in a vacuum and speak with voices devoid of humanity.

## Prompting the Dramaturgical Conception

The fourth and final stage of the generation process involved the integration of the theme, setting and character into a coherent dramaturgical structure. This could be either a three-act structure (for shorter formats) or an eight-sequence structure (for longer formats). The process began with an outline, proceeded to the creation of summaries for each scene (breakdowns) and then involved fleshing out each scene with details.

(1) Examples for outlining the whole film: '*create a story outline for the moral meaning 'absurdity of competitiveness'. include the character description, recall the conflict, the protagonist, the antagonist, the backstory of each of the characters described.*'

This resulted in a title, genre, setting, an analysis of themes included in the moral meaning, plot outline with introduction, rising action, conflict, climax, falling action and resolution.

(2) Examples for the scene breakdown: '*create a list of scenes, based on your outline. Every scene has an objective, and a summary. Include who is in each scene. Analyse the main conflict, make sure it is included in the scene.*'

This resulted in an extensive list of scenes:

Example-output:

*Scene 1: Introduction of the Ice Bubble: Present the first phase of the Ice Bubble's journey, symbolising natural beauty and the beginning of a journey.*

*Scene 2: The Vodka-Driven Atmosphere: Show the contrast between the lively, vodka-soaked atmosphere of the ship and the focused demeanour of Kata, Anna, and Lena.*

*Scene 3: The First Meeting and Challenge: Illustrate the tension between Alexei's dismissal of women's soccer and the protagonists' determination.*

*Scene 4: Sealing the Bet with Vodka: Show the cultural and personal significance of the bet being sealed with vodka, highlighting the hesitance and eventual participation of the women.*

*Scene 5: Introduction of the Ice Bubble – Phase Two: Depict the second phase of the bubble's journey, symbolising tension and progression.*

*Scene 6: The Climax – The Game Begins: Capture the fierce competition on the ice, emphasising the physical and emotional stakes.*

*Scene 7: The Climax Continues – Anna's Fall: Highlight the dramatic fall of Anna and the consequential decision by her teammates to continue the game.*

*Scene 8: Falling Action – The Aftermath and Diagnosis: Show the immediate aftermath of the game and the devastating news delivered to Anna.*

*Scene 9: Conclusion – Anna's Defiance: Depict Anna's personal rebellion against her fate and the uncertain future.*

The later we progressed in the film, the more often GPT had to be reminded of the main conflict and of integrating the characters' backstory into the narrative. Due to the lack of planning-head, prompts like '*do not solve the conflict, leave it open until the end*' were necessary very often to avoid a premature ending of the narrative.

(3) Example of individual scene writing: '*Write scene no. 4 from the list of scenes. Tell us visually how the story on the ship continues with the women and Alexei. Don't use 'CUT TO' or 'FADE OUT'. Be playful, direct, use visual language, like in an industry standard screenplay format. Do not include interpretations. Take into account what happened before, creating a convincing chain of events. Avoid stereotypes. Use surprise and emotions and human interactions and dynamics. Do not write dialogue, only summarise or paraphrase what the characters are expressing verbally.*

This resulted in the generation of a complete scene in screenplay format. The decision not to include dialogue in the generated text is based on the observation that GPT-4 is not well-suited to this task. The model struggles to write dialogue that is coherent with the narrative (see step 'Character Conception'), to link dialogue to the narrative, and to use emotion instead of information and to create subtext.

The Four Stage Story Generation with its focus on theme, story, character and dramaturgy was repeated multiple times, then analysed and re-prompted, curated by humans. The most difficult part remains the transition between the four different stages of story generation. Frequently, results were not transferable from one stage to the other, as for example the character's backstory did not sufficiently influence the dramaturgical conception. Working systematically with step-outlines facilitated that the dramaturgical conception could be geared towards a systematic ideation and story development process.

### The Parrot on Polar Ice

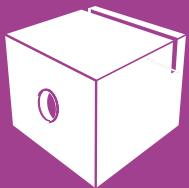
The final result of our Four Stage Story Generation experiment, the screenplay *Frost Bite*, demonstrates that despite a prompting strategy, the human intervention factor is still a decisive element in artistic screenwriting. The weaknesses are too prevalent for submissions to the screen or to screenwriting competitions or use as a basis for professional filmmaking. We will be aiming at a significant rewrite in a 'classic' co-creation (among human actors). The principle weaknesses of *Frost Bite* are that the thematic conception is still too weak (the AI could not generate a satisfactory story on the complexities of competition), while the story world conception is convincing and definitely adds a captivating layer to the story, the character conception lacks human depth (notably in the absence of engaging inner conflicts and the schematic and predictable character design) and the dramaturgical conception shows surprising flaws in some scenes (especially the dramaturgy of the soccer game – rather standard sports dramaturgy that can be

found daily e.g. in sports reporting – lacks suspense and coherence). Even in a human-AI collaboration, the dramaturgy lacks what Cowgill (1999) describes as 'a certain amount of surprise and excitement while at the same time maintaining momentum and meaning'. While GPT certainly has the potential to maintain momentum, in the context of artistic screenwriting it fails to develop a consistent meaning. Generative AI is non-deterministic; it does not have a perception of reality nor common sense. It does not 'know' anything. Its 'knowledge' is limited to training data. Human intervention cannot be dispensed in automated writing at this point, yet. Despite its limitations, in future screenwriting projects we are aiming to apply and refine the Four Stage Story Generation as a development method. But one of the essential findings is that, at this point, the fifth stage of the story generation is the stage of 'human rewriting'.

Our findings thus confirm that LLMs are a 'stochastic parrot' (Bender 2021), producing token after token without planning-ahead and based on ever-growing training data. Millière (2022) claims that 'sufficiently advanced mimicry is virtually indistinguishable from intelligent behaviour'. But AI mimicry in storytelling without a deeper understanding of human complexity is not (yet) capable of Aristotelian mimesis. The stochastic parrot in the writers' room is a penned parrot. In the absence of embodied experience, mortality and any stakes in human existence, LLMs, thus far, are limited to the role of a penned parrot. This penned parrot is caged by the abstract machine algorithms and therefore distanced from the human condition. Confronted with too complex situations in a non-native habitat such as the polar ice, the parrot is (to date) highly dependent on human assistance.

## Bibliography

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Bolt, B. (2006). A Non Standard Deviation: Handlability, practical knowledge and practice led research. Speculation and Innovation: Applying practice led research in the creative industries.
- Cowgill, L. J. (1999). *Secrets of screenplay structure*. Lone Eagle Publishing Company.
- de Assis, P., & D'Enrico, L. (2019). *Artistic research-charting a field in expansion*. Rowman & Littlefield.
- Deuerling, T. (2016). *Innovationsmanagement für neues Fernsehen: Entwicklung von Bewegtbildformaten in Abhängigkeit vom Innovationsgrad*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-11669-9>
- Dunne, W. (2009). *The dramatic writer's companion*. The University of Chicago Press.
- Field, S. (2005). *Screenplay: The foundations of screenwriting*. Bantam Dell.
- Hendry, M., Kottmann, N., Fröhlich, M., Bruggisser, F., Quandt, M., Speziali, S., Huber, V., Salter, C. (2023). Are You Talking to Me? A Case Study in Emotional Human-Machine Interaction. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 19, no. 1, October, pp. 417–24., <https://doi.org/10.1609/aiide.v19i1.27538>
- Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D. d. L., Hendricks, L. A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican, K., Driessche, G. v. d., Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., & Sifre, L. (2022). *Training computer-optimal large language models*. <https://arxiv.org/abs/2203.15556>
- Howard, D., & Mabley, E. (1993). *The tools of screenwriting*. St. Martin's Griffin.
- McKee, R. (1997). *Story*. Methuen.
- Millière, R. (2022). *Moving beyond mimicry in artificial intelligence*. Nautilus. <https://nautilus.us/moving-beyond-mimicry-in-artificial-intelligence-238504/>
- Nelson, F., Levy, O., Sap, M., Tenney, I., Kuhn, J., Reif, E., Zheng, C., Hou, Y., Pavlick, E., & Bowman, S. R. (2023). *Lost in the middle: How language models use long contexts*. arXiv. <https://doi.org/10.48550/arXiv.2307.03172>
- Riedl, M. O., & Young, R. M. (2010). Narrative planning: Balancing plot and character. *Journal of Artificial Intelligence Research*, 39, 217–268. <https://doi.org/10.1613/jair.2989>
- Russin, R., & Downs, W. (2012). *Screenplay: Writing the picture*. Silman-James Press.
- Sharples, M., and Pérez y Pérez, R. (2022). *Story Machines: How Computers Have Become Creative Writers*. 1st ed., Routledge, <https://doi.org/10.4324/9781003161431>
- Sorkin, A. (2020). *Aaron Sorkin teaches screenwriting* [Video]. MasterClass. <https://www.masterclass.com/classes/aaron-sorkin-teaches-screenwriting>
- Trottier, D. (2014). *The screenwriter's bible*. Silman-James Press.



EARLY VISUAL MEDIA LAB

C I C A N T