# The AI Director: From Document to Documentary

**David A. Noever**
**PeopleTec, Inc.**
**Huntsville, AL**
**david.noever@peopletec.com**

**J. Wesley Regian**
**PeopleTec, Inc.**
**Huntsville, AL**
**wes.regian@peopletec.com**

## ABSTRACT

For gaming, training, and video directing, the high cost and intensive labor requirements limit the transformation of the initial abstract concept to the final product, either as a film or as interactive software. This paper explores the steps required to translate rough instructions into instructional materials. Where possible, we insert machine learning to automate those transformative steps from script to movie. The DoD repositories of archived training manuals, reports, and lessons learned include a vast and varied mixture of stepwise instructions and industrial-style line drawings. We model the AI input as a human-machine interchange, augmenting the role of the human *auteur* with an automated associate director. The example training builds on the starting material of an Army field manual. Still, it is supplemented in our case by the modern AI tools to script a set of original instructions using natural language processing. We apply custom text-to-speech tools to dub audio tracks and narrate the required assembly steps. Using the visually creative elements of generative adversarial networks, we complete the required visual representations from the initial script as single images, then animate reels and entire vignettes for video storytelling. To evaluate the output quality, we compare introductory training material made entirely by human directors against the machine-augmented version. The original training material walks soldiers through an example Army field manual (FM2.22.3), *Human Intelligence Collector Operations*. Quantitative metrics are applied to evaluate the instructional effectiveness of source material and compared to AI-generated training videos based on knowledge test items. Critical knowledge elements and knowledge test items are automatically generated during the development process. The work assesses the training video quality using expert ratings from trained 35M military intelligence specialists.

## ABOUT THE AUTHORS

**David Noever** has 27 years of research experience with NASA and the Department of Defense in machine learning and data mining. He received his Ph.D. from Oxford University, as a Rhodes Scholar, in theoretical physics and B.Sc. from Princeton University, summa cum laude and Phi Beta Kappa. While at NASA, he was named 1998 Discover Magazine's "Inventor of the Year" for the novel development of computational biology software and internet search robots, culminating in co-founding the startup company cited by Nature Biotechnology as first in its technology class. He has authored more than 100 peer-reviewed scientific research articles and book chapters. He also received the Silver Medal of the Royal Society, London, and is a former Chevron Scholar, San Francisco. His primary research centers on machine learning, algorithms, and data mining for analytics, intelligence, and novel metric generation.

**J. Wesley Regian** has 32 years of experience in cognitive performance modeling and knowledge-based software technology development, primarily for military applications with AFRL, AFOSR, USASMDC, and DARPA. His work has supported over 50 fielded systems. He has published over 100 papers on psychometrics, skill acquisition, artificial intelligence, simulation-based training, computer-based training, intelligent computer-based training, virtual reality for training, intelligence analysis, human terrain modeling, knowledge representation, knowledge management, human learning and memory, individual and developmental differences in human cognition, spatial ability, and spatial information processing, cognitive modeling, and cognitive automaticity. Dr. Regian was a National Research Council research adviser for ten years while Senior Scientist for Knowledge-Based Systems at the US Air Force Armstrong Research Laboratory.

# The AI Director: From Document to Documentary

**David A. Noever**
**PeopleTec, Inc.**
**Huntsville, AL**
**david.noever@peopletec.com**

**J. Wesley Regian**
**PeopleTec, Inc.**
**Huntsville, AL**
**wes.regian@peopletec.com**

Keywords: *Authoring tools, Automation, Content Generation, E-Learning Applications, Emerging Technologies*

## INTRODUCTION

It is famously expensive and time-consuming to develop good-quality e-learning applications. We are developing capabilities for applying current generation Artificial Intelligence and Machine Learning (AI/ML) to automate and thus significantly reduce the expense and time required to generate good quality training applications. Here we describe initial progress on what we consider low-hanging fruit in this technical endeavor. We do not aim to automatically develop Intelligent Tutoring Systems, Simulation-Based Training, or highly interactive computer-based training. Our less ambitious initial goal is to largely automate the creation of good quality military training videos using textual (including graphics, tables, lists, etc. - see Figure 1) knowledge artifacts (military field manuals) as source material. Training videos alone can lead to significantly better learning outcomes than instructional manuals and textual documents. Improved median outcomes for video-watching students can exceed one to two grade levels compared to learning from textbooks (Stockwell et al., 2015). Our AI/ML approach has an additional pedagogical benefit beyond just automating content creation to reducing video development costs. By using AI to derive knowledge objects linked to instructional objectives, instructional test items, and presentation content – we get mastery learning capabilities for free in the resulting application. In the simplest form, mastery learning is taking steps to verify that students have learned (mastered) prerequisite material before proceeding to a higher level, more derivative instructional objectives. Mastery learning techniques can routinely enhance learning outcomes by one standard deviation (Bloom, 1968). Bloom's finding has been replicated many times in multiple instructional contexts.

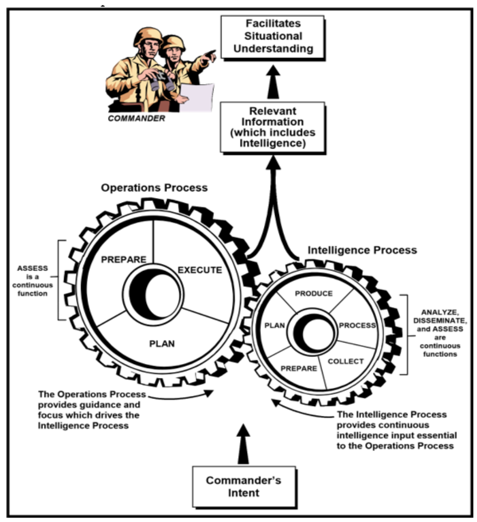


**Figure 1. Example Operational View of Human Intelligence Collection. The logic map seeks to summarize an entire field manual in a single core graphic.**

This research explores the current limitations of automated and original digital instructional content creation. Previous attempts to convert military documentation into training material draw heavily from the simulation and game industry, focusing on generating vignettes to illustrate acquired skills. Artificial intelligence (AI) or machine learning (ML) offers new approaches to consider beyond just more realistic game avatars or question and answer cues. To illustrate the automation potential, one can pose the AI project-defining question, "Describe the key steps in making a good military training video?" As described in Methods, an original text generation algorithm (Brown et al. 2020) yields a beneficial five-step formula or method: 1) define your purpose and audience; 2) research your topic; 3) write a script; 4) choose your visuals; 5) edit your video. As implied by the paper's subtitle, "From Document to Documentary," the first three steps are essentially encapsulated within a pre-existing military field manual, a rudimentary narrative ("Document") which the AI methodology will convert into filmable artifacts ("Documentary"). In his paper, "Video Killed the Textbook Star," Rackaway (2012) found multimedia "improve student learning outcomes generally and specifically show the greatest improvement in written test components, especially with students who struggle early in a course."

The research first seeks to convert unstructured text to knowledge, extracting the relevant storytelling elements (facts, assertions, procedures) and entities (persons, places, organizations), modeling key topics, language translation or optical character recognition where needed, and finally, script generation. Once these preliminary models assemble a draft AI-driven script, multimedia creation begins with character generators and animators. The final editing requires human input as presently operating but one envisions automating a rules-based script to patch together video and audio layers. The original contributions of this approach center around first the deconstruction of AI-augmented steps needed for completing draft training material, followed by metrics to evaluate key story points or scenes where human directors need manual editing and new content additions. The results span the AI creation of original digital elements and their real-world building blocks for matching task instructions to successful completion.


**Figure 2. Word Cloud Created to Capture Main Concepts of Field Manual**

**The Training Problem: Script and Video**

Automated instruction refers to computer-assisted teaching, computer-based training, intelligent tutoring, simulator-based, interactive video-based, part-task, and embedded training (Regian and Shute, 1992, 2013). As an anecdotal evaluation, soldiers describe the DoD's typical instructional approach for its Interactive Electronic Technical Manual (IETM) as a "Great concept, but poor implementation" (Steward, 2004). A standard critique of field manuals (FM) is that while they are readily available in battalion libraries, the field manual is not used in the field. The soldier is unlikely "to carry it around…chances are they'd only consult it when something unusual came up (Engber, 2005). The training never gets fielded.

The challenge of creating compelling content suffers from the need to be original yet relevant to the curriculum. In its simplest form, a training video involves a minimal narrative script and video production. What elements of these two can be automated or at least augmented by machine-human collaboration? The field manual might be an initial digital piece of narrative content or the training film's initial script. In an essay entitled "Broken and Unreadable: Our Unbearable Aversion to Doctrine," West Point's Modern War Institute (MWI) promotes an instructional view that manuals should offer doctrine that is "both compelling reading and interesting…rich as it is deep" (MWI, Leonard, 2017). The MWI essay highlights numerous shortcomings for future improvements: reducing the sheer number of manuals, shortening their reading length, untangling their intertwined subject matter dispersed across multiple volumes, minimizing their tendency to over-define terminology, and removing the distortion of commonly used words (like "asymmetric") as used repetitively as specialized military jargon. The author echoes the soldier's quip, "it's only a lot of reading if you do it."

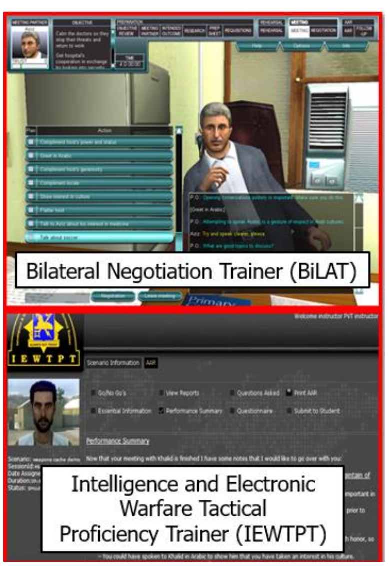


**Figure 3. DoD Training Scenario Software.**

When a potential solution emerges to edit the manual and fix these flaws, the final MWI assessment stressed that the Army manual transforms into contingent and intermediate guidance, the opposite of its intended role as final guidance. The library of manuals is nearly always in flux, despite touting their primary reason to exist as a reference wholly centered on the few core and fundamental principles. As a teaching tool, the manual's "writing isn't always the best…lower than your average comic book."

In addition to the text script, what visual content do the field manuals provide? The MWI goes on to savage the visual content, the field manual's core graphic, and a logic map or operational concept that walks the reader through the manual's big ideas (Figure 1). Not only does this training approach reward soldiers who skim the words (in favor of viewing one summary picture), this heuristic produces instructional outcomes that are only "PowerPoint deep" or "logic map deep." The learning stops at the operational view (OV1) stage. The present research explores an augmented role for machine learning and artificial intelligence in converting written field manuals into future production components. As an alternative starting point, an augmented visual map features a concept or word cloud (Figure 2), which shows the relative significance of HUMINT trade terms using the Wikipedia description (Army Field Manual, 2006).


**Figure 4. Cartoon Narrator for Clearance.Jobs (2021)**

The DoD has long invested in artificial intelligence (AI) (see Regian et al., 1989). This effort has explored advanced text-to-speech for more realistic voicing (Noever et al., 2021) and introduced new characters, avatars, or automated agents in a training simulator (e.g., flight instruction) or wargame (Ablanedo, 2018; Parodi, 2013). Figure 3 shows two popular training software packages with similar aims to instruct deployed soldiers on interactions with the local population. The Bilateral Negotiation Trainer (BiLAT) offers an interactive, game-like simulation of negotiating in different cultures (Kim et al., 2009). The Intelligence and Electronic Warfare Tactical Proficiency Trainer (IEWTPT) also uses simulated and live scenario data to train intelligence officers (Williams et al., 2002). A previous study (Noever & Regian., 2019) applied machine learning to automate negotiating game scripts for training Army negotiators. The intelligent tutor in these applications serves as a guide to case scenarios and live role-play exercises, but "when this instruction is available, it tends to be limited by its instructional capability, cost, or both" (Kim et al., 2009). Figure 4 shows a cartoon-like narrator describing the HUMINT collection process (Clearancejobs, 2021). The reliance on stick-figure avatars (Ablanedo et al., 2018) as training instructors provides "a low-cost real-time puppet character in a virtual environment."

One plausible reason these training frameworks generally do not rely on (or benefit from) AI automation has been the technical immaturity of any creative machine learning tools. The special effects of game and movie designers rely on expensive specialists working the stages from concept to editing. However, growing areas for these automated AI methods can assist instructors with document navigation, search, branching narratives based on student mastery, and expert systems of "if-then" rules. A practical, interactive lesson typically involves non-linear storytelling and narrative tags now common to online documents such as XML, HTML, and SGML markup languages that blend text, graphics, databases, and multimedia video and audio. AI tools (Table 1) seek to reduce longer text sections into filmable script scenes, generate voice-over and background audio, and animate teachers and students into a realistic role-playing exercise to accelerate the time-consuming creation of digital content. From the instructor's view, a computerized expert system improves productivity and accelerates multimedia output. From the student's view, automated instruction disseminates more types of media at a lower production cost, thus matching teaching methods to individual needs or preferences. Where possible, the research will test the strength of AI-as-a-Service (AiaaS), which connects multiple algorithms into a single production pipeline using Application Programming Interfaces (API). Most cloud services offer free and paid versions of large AI models that leverage Graphical Processing Units (GPU) banks to accelerate video and audio production.

## METHODS

### Army Field Manual.

US Army Training and Doctrine Command (TRADOC) includes 61 current field manuals online (APD, 2022), 8846 Technical Manuals (TM), and 463 Graphic Training Aids (GTA). Older repositories and libraries indexed over 650 Army field manuals (FM)(Global Security, 2015) before the DoD limitation to around 50 core FMs (Doctrine 2015) and supplemental or lower-level series called Army Techniques Publications (ATP). This study selected the Army Field Manual titled "*Human Intelligence Collector Operations*," FM2-22.3, or HUMINT collection (2006). As a measure of its relative influence, the Intelligence Field Manuals (FM-2) rank 29th among the top 50 web views (Army Pubs, 2022), or around the median between the most popular FM-6 series on Commander and Staff Operations (2014) and the least popular FM-4 Sustainment (2019). Since 2009, authorized users can update the manuals similar to an Army living document or "wiki" (Cohen, 2009).

The 384-page PDF manual for HUMINT teaches the operating procedures for collecting human intelligence in four sections and thirteen chapters, ranging from analysis tools to effective communications. One motivation for selecting this HUMINT Army manual stems from the possible comparisons between parallel human and machine intelligence, mainly the instructions on evaluating an adversary's capabilities, vulnerabilities, and intentions. The manual describes how soldiers gather insights by questioning, screening candidates, interrogating them, and debriefing their command. The manual explains nineteen interrogation methods ranging from rapport building to incentives, inducements, and rewards for eliciting information from detainees in US custody. The goal is to convert this document into video or documentary materials. Notable machine-ready tasks include basic document exploitation (DOCEX) and pattern analysis with link diagrams. Table 1 summarizes a pipeline for document exploitation designed to extract metrics and foster new narrative scoring methods from the HUMINT collection guide.

**Natural Language Processing (NLP).** The research employs the most extensive natural language model from OpenAI (licensed to Microsoft). The model is trained on a text corpus representing approximately 10% of the internet and uses a machine learning model architecture called transformers (Generative Pre-trained Transformer-3). The transformer approach has found success in modeling the long-term correlation between linguistic concepts (e.g., "queen-woman, " king-man," etc.). Unlike other natural language tasks such as chat-bots, the main application of GPT-3 highlights its creative capabilities to automate up to two pages of the original text. The narrative output is verifiably original by a direct test against Google search and online plagiarism detectors. The usefulness of creating training videos stems from the model's ability to respond appropriately to prompts or questions such as "what is the job of a human intelligence collector?" By concatenating multiple responses, an entire training scenario or script can be constructed in an automated fashion using the journalistic approach of asking the 5W1H format (e.g., who, what, where, when, why, how?).

As a concrete example, we asked the model to describe the critical parts of compelling military training videos as a test. These five hints, all auto-generated from a linguistics model, guide the research goals.
The model responds with the following advice and guidance.

*There are a few key ways to train soldiers effectively using video:*

1. *Make sure the video is high-quality and clear.* This will help soldiers to see and understand the information better.
2. *Use real-life footage whenever possible.* This will help soldiers see how the teaching techniques can be applied in a real-world setting.
3. *Use voice-over narration to explain the information in the video.* This will help soldiers understand the information better and keep them engaged.
4. *Use on-screen text to highlight key points.* This will help soldiers remember the information better and help them follow along with the video.
5. *Keep the video short and to the point.* This will help soldiers stay focused and prevent them from becoming bored.

The metrics used to score training success pose a challenge. While the research recommendations vary, our study simply posed the question to GPT-3, how best to grade training outcomes? The response summarizes multiple approaches "The metrics to use to assess how effective a soldier is learning from a training video would include how often the soldier watches the video, how much time the soldier spends watching the video, how many times the soldier pauses the video, how much time the soldier spends rewatching sections of the video, and how many notes the soldier takes while watching the video." This machine-generated response was tested as original content using the online originality meters and Grammarly syntax checker. The first answer, however, is more observational, as behavioral grades, rather than demonstrating specific proficiency before and after the short training session.

It is worth noting that the NLP model crafts its response more as a universal linguistic interpreter without specific training in any specializations (so-called "zero-shot or few-shot learners" Brown, 2020). In other words, users of the model need not provide extensive language examples before asking a question in a specific field, even in technical areas such as medicine, law, or engineering. The user prompt, however, does guide the expected response, both in narrative tone (e.g., instructional) or content (e.g., specialized jargon). To tease out more specific measurement criteria for our training script, we posed a more specific prompt for learning evaluation (e.g., "Be as specific as possible") and received a more varied spectrum of scoring approaches.

*The best ways to score how much actual learning has been transmitted from a training manual to a soldier in practice are to:*

1. *Give a pre-test to soldiers before they begin training.* This will serve as a baseline measure of their knowledge.
2. *Have soldiers complete the training manual and then take a post-test.* This will show how much they have learned from the manual.

3. *Have soldiers complete a practice exercise using the skills they have learned from the manual.* This will show how well they can apply what they have learned.
4. *Observe soldiers as they complete a task or mission related to the training manual.* This will show how well they can use the skills they have learned in a real-world setting."

**Figure 5. Realistic Avatars for Narrating Training.** ***These actors verifiably do not exist and never have historically in a copyright sense.***

To guide the model to answer questions about human intelligence collection, we prefaced the prompts with a joint contextual statement, "*You are an expert on collecting human intelligence for the US Army.*" This stray contextual cue or compliment proved essential to getting quality responses by injecting a primary narrative thread in which the model could establish credible sequencing and tell a coherent story based on specific questions a student might pose. By paraphrasing the field manual's table of contents as crucial topics for the training video to address, we generated a 3-page script or a 9-minute instruction covering the 13 chapters and nearly 400 pages of raw material. Following the advice to include screen text, we overlay essential student questions as captions and voice-over narrators reading the training script.

**Video and Multimedia Generation.**
The strategy to bypass expensive actor and voice-over costs rely on the creative AI field called Generative Adversarial Networks (GANs). In short, two trained algorithms compete against each other, one spawns fake images, and the other edits them by comparing the outputs to the desired goals. This method creates fake actors (Mansourifar, 2020), who then can be animated to narrate the AI-generated script. We selected actors (examples shown in Fig 5) from an online persona-generating service (thispersondoesnotexist.com) based on training against 1000s of models and choosing intelligence agents of different ages, genders, and career stages.

**"Deep Fake" Video Generation.**
The production value of still images is limited in storytelling. The recent genesis of "deep fake" video technology now enables the animation of still images as potential storytellers, each with their dubbed voices and facial mannerisms. This technology began with low-resolution face-swapping but has progressed to convincing creations as both script-readers and persuasive actors. Using a GAN, the online biography service MyHeritage "Deep Story" offers a convenient machine learning platform for generating multiple avatars as test script readers (MyHeritage.com, 2022). For the HUMINT collector series demonstration, we animated the avatars or "persons who do not exist," along with one historical figure from World War II: Ian Fleming's James Bond character, Dusko Popov—referred to as the world's greatest spy and MI6 double agent. Historians credit Popov with assisting the D-Day Invasion after diverting the Germans from defending Normandy and convincing his Axis-power handlers that the Allies would instead land in Calais. Popov's video segment also serves as an avatar or synthetic example created from a single still image, which subsequently animates as a spokesperson to describe his real-life spy-craft and intelligence-gathering prowess.

The tedious editing of video vignettes proved to be the least automated portion of the experiment. To join the avatars as training instructors, we chose the CapCut video suite running inside an Android smartphone emulator and PC desktop using BlueStacks (Minor, 2021). While less functional than complete editing suites, fast online video production's speed and cost savings justified its use for editing less than a 15-minute training short. Selected special effects included other machine learning modules. These specialized algorithms remove image backgrounds surrounding faces (automated "green screening techniques"). They zoom transitions between scenes (sometimes called the "Ken Burns 3D effect" in honor of the award-winning documentary director who employs it to animate historical still images from the Civil War).

**RESULTS**

The study results involve the semi-automated generation of an example training video given only the dense source material of an Army Field Manual. The output is shown online as a 13-minute instructional video (Noever and Regian, 2022). The method of evaluating the video's effectiveness centers on an expert survey of a (35MOS-qualified) former intelligence operator and pre-/post-testing for non-expert human subjects.

**Findings of the Field Manual Evaluation**

As a baseline to quantify the effectiveness of the HUMINT field manual as training material, the research applies narrative metrics for comprehension and complexity. The graphic content includes line drawings and tables sprinkled across the chapters, with an average of one graphic per 8-10 pages. As a source for video production, the charts, tables, and hand-drawn illustrations yield little usable digital content. An AI-driven grammar scan (such as Grammarly or Microsoft Spell Check) reveals more than 100 errors, including non-American spelling choices ("honour" and "endeavour"), along with confusing usages that typically evade mechanical spell checks (such as using "past" as a verb, not "passed"). Of particular concern with a finished editorial product to train human interrogators, the manual mixes the verb "polices" when the actual noun is "policies" as about *AR 381-143. Military Intelligence Nonstandard Material Polices (sic) and Procedures, Bibliography 2*). Following the expert writing tips from Grammarly Premium, 5920 writing issues (or 15 issues per written page and 3% of total sentence content) arise based on poor scores in engagement, delivery, correctness, and clarity. The Grammarly model flags more than 100 examples each for categories of word choice, misuse of passive voice (22.9% of total), wordy or unclear sentence structure, and compound punctuation. The point of aggregating these metrics is not to nit-pick its style choices but to highlight the opportunities available for machine learning models to correct long narratives with automated editorial assistance. Compared to its database of other writers, Grammarly's AI model scores a representative Army manual page 83/100 for readability (word and sentence length) for college graduates but with a below-average rating for uncommon words (39%). From a trainer's perspective, the manual rates "mostly clear" content but "a bit bland" for engagement with examples like repeating "normally" three times in three connected sentences.

**Figure 6. Example Video Sequence to Narrate Script Delivery. The CapCut editor aligns the audio, video, and transition sections.**

**Pedagogical Field Manual Example in Text Analysis**

The HUMINT manual (FM2-22.3) requires the soldier to read at a high school literacy level (Flesch-Kincaid Grade Level, 12.4). The standard field manual is supposed to be written at the sixth-grade level (Engber, 2005) or "hover at the eighth-grade level (Leonard, 2017). Because grade level scores use syllable counts and sentence length, they underestimate complexity when the text includes acronyms. The HUMINT manual lists more than 300 unique acronyms, containing 1% of its total word count as acronyms (7118). While not atypical as a style for such guides, these 18 acronyms per page require the soldier to understand technical jargon and remember complex terms across long narrative chapters and sections. If a soldier comprehends as an average reader, they could finish the manual in 10.7 hours in a single session at 300 words per minute (WPM) or about one page every two minutes (SwiftRead, 2022). If a production team chose instead to convert the existing field manual into an audiobook, the studio time to read it aloud (175 words per minute) would span 18-19 hours or approximately two long working days. To create the equivalent voice-over with an actor in a studio would exceed $5000 for a two-day rental, essential equipment, and two crew members (Open Media Foundation, 2022). Without including copyright and residual fees, published estimates for hiring non-celebrity actors quote up to $400 per hour (Winograd, 2022). If one imagines the extreme case of having six actors (Figure 5), the multi-day cost of the production cast exceeds $45,000. An alternative experiment using Microsoft's automated Text-to-Speech (TTS) tools (Read Aloud) converted one page every 3 minutes as an unattended audiobook creator. As automated production, a sample page is available online (Noever, 2022).

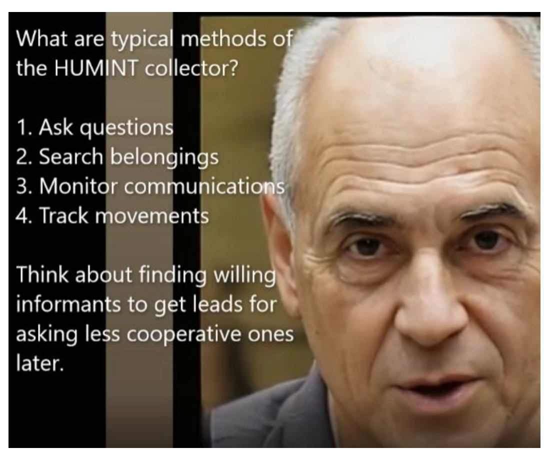


**Figure 7. Scene from automatically written and filmed instructions**

**Pedagogical Short Script Examples in Text Analysis.**
The main goal of creating a 10 to a 15-minute instructional video that summarizes a field manual relies on the 3-page script with a 5-minute reading time and 9.3-minute speaking time by animated tutors. Following the expert writing tips from Grammarly Premium, two writing issues in the GPT-3 automated script (or 0.6 points per written page and 0.1% of total sentence content) arise based on poor scores in engagement, delivery, correctness, and clarity. The Grammarly model identifies no AI-written sentences that are incorrect. The model does propose structural changes to increase transparency ("A bit unclear" rating), improve engagement ("A bit bland") based on word choices, and enhance delivery ("Slightly off" rating) based on syntax conventions. Interestingly the Grammarly model rates the entire narrative script as original content with no plagiarized sentences. Readability statistics for the generated script (Flesch-Kincaid Grade Level, 11.9) do not differ dramatically from the field manual itself (12.4). The automated narrative offers a higher readability score (36) than the field manual (32.8).

To increase the overall readability of the script, we requested via the OpenAI API a summary of the entire field manual script (Grade Level 2), including a description of the job of a HUMINT collector:

> *"In what ways can you influence the behavior of the interrogated person? There are a few ways to influence the behavior of a person. The first is to establish rapport. This can be done by asking about their family, their work, their hobbies, etc. The second is to be respectful and humble. Show that you are interested in what they have to say and that you value their opinion. The third is to ask open-ended questions. This will encourage the person to open up and provide more information. The fourth is to listen carefully to the answers. Pay attention to body language and other nonverbal cues. The fifth is to follow up with additional questions. If something doesn't make sense, or if you want to know more about something, don't be afraid to ask."*

**Scripting an Automated Training Video**.
The complete video is available for review online (Noever and Regian, 2022) and illustrated for narrative sequences in Figure 6. The characters voice the narration with cut-scene transitions showing intelligence collection vignettes. The overall engagement, clarity, and correctness of the training script compares favorably to the human production of the Army field manual. Pre- and post-training surveys suggest that the AI video communicates the key concepts in a testable format that, for larger population samples, could, in principle, outscore the "logic map" approach of single graphics in this instructional genre. The generation of narrator avatars and their scripted voice animations represent key AI contributions to building the 13-minute film. Scene transitions and sound editing complete the initial draft of the HUMINT short documentary.

**Table 1. Pre- and Post-Test Example Generated to Score Video Training Effectiveness**

| Question | Answer |
|---|---|
| What is the most important nonverbal cue? | Body language |
| What are the limitations of HUMINT collection methods? | A large amount of manpower, the need for a high level of training, the possibility of human error, and the potential for bribery and corruption |
| Is developing informants a HUMINT capability? | Yes, it is a HUMINT capability |
| What can a HUMINT collector withdraw if the person being questioned is not providing useful information? | Incentives |
| Name some reasons to offer incentives in HUMINT collection? | Provide more information than they would otherwise, be more forthcoming and truthful in their answers, and make the person more cooperative overall |
| Do incentives encourage truthful answers? | Yes, incentives encourage truthful answers. |
| What do HUMINT collectors develop with their source? | Relationships |
| What is the difference between police and HUMINT collectors? | Police enforce physical security. HUMINT collectors gather intelligence from local sources. |

| | |
|---|---|
| What are the key elements for a good HUMINT collector? | Understanding local culture, customs, and language |
| Do HUMINT collectors learn local languages? | Yes, HUMINT collectors learn local languages. |
| What kind of sources provides valuable intelligence? | Local residents |
| How does one get truthful responses? | Establish rapport. Ask about family, hobbies, work |
| Do HUMINT collectors ask open-ended questions? | Yes, HUMINT collectors ask open-ended questions. |
| What methods are banned by the Geneva convention? | Torture, threats, coercion, deception |
| Do HUMINT collectors follow Geneva Convention? | Yes, HUMINT collectors follow the Geneva convention. |
| Name a reason to remove incentives? | The collector may withdraw the incentives if the person being questioned is not providing useful information, if they are being evasive, or if they are not being cooperative. |
| Do people become more cooperative as incentives are withdrawn? | Yes, the collector may withdraw the incentives if the person being questioned is not providing useful information, if they are being evasive, or if they are not being cooperative. |

Compared with YouTube shorts (produced by Army Recruiting, National Guard, Clearance.Jobs) and other sources (YouTube, 2022), the machine examples provide comparable human-generated content. Still, the 1–3-minute formats do not attempt to cover all the fundamental concepts of the 400-page field manual. Table 1 summarizes initial expert questionaire and sample scores in a pre-/post-test quiz.

**Test Development for Scoring the Video**.
The testing protocol benefited from the deep learning platform called QuestGen (Golla, 2022). This open-source project takes its input as large text volumes and generates an automated quiz of multiple-choice or binary (true-false) questions. Based on selecting 18 general questions, we asked non-experts on human intelligence to take a pre-and post-test, then evaluated the effectiveness by comparing acquired knowledge. Table 1 summarizes the quiz construction to understand the training effectiveness. Given the small sample size, a more informative evaluation included a simple survey of knowledgeable or certified experts, these anecdotal comments guided some additional improvements during editing. The two evaluators noted that the video avatars could be distracting because of mechanical voicing, spurious lip or eye movements, and the overall "bobbing head" effect while narrating. Both reviewers cited the lack of more extended vignettes longer than thirty-second scene cuts, even when scripted together into longer narratives greater than ten minutes.

**Scalability of AI Video Production Methods**
To assess the mass production value of converting Army field manuals into automated tutorial videos, we turned two other manuals into online demonstrators (Figure 8, and Noever, Regian, 2022). With a total production time of thirty minutes per video and no script, actor, or copyright cost, this approach supplemented the HUMINT Field Manual (FM 2-22.3) exemplar with additional candidates available for scoring and peer review (i.e., Cyber (CEMA) Field Manual FM 3-12 and Chemical, Biological, Radiological, And Nuclear Operations, CBRNE Field Manual FM 3-11). When

tallied together, the project generated thirty-three minutes of total video, which commercial costs otherwise would have exceeded a half-million dollars.


**Figure 8. Three AI Training Videos Demonstrating Automated Conversion of Army Field Manuals Without Actors, Script, or Voice Production.**

## CONCLUSIONS AND FUTURE WORK

The research explores the capabilities and limits for automatically generating text, audio, and video content for training instruction. The text content extracts a filmable narrative script and derives its critical features from a "few-shot learning expert" as its author. The film production similarly derives its avatars as realistic but fictitious actor profiles (including generated headshots), which subsequently animate as audio and video "deep fakes." The resulting combination provides a draft for human editing that stitches the image frame sequences into a coherent teaching video. Future work should explore alternative AI customizations outside the large AIaaS and API, so that subject matter expertise becomes more specific and relevant to the students. Additional work is needed to quantify the testing results based on broader sample populations and more in-depth surveys of HUMINT 35M-certified soldiers with field experience and example training history. Further work needs to include better scoring methodologies, particularly those that offer cohorts who may read the field guide alone, read and watch the video, or pursue periodic evaluation and problem-solving during vignettes. Where possible, the research points out the cost and time advantages of augmenting film production without employing studio time, copywriters, actors, or research staff, an outcome that speaks to the potential for AI-enhanced versions of all major Army and DoD-wide training documents.

## ACKNOWLEDGEMENTS

The authors would like to thank the PeopleTec Technical Fellows program for its encouragement and project assistance.

## REFERENCES

Ablanedo, J., Fairchild, E., Griffith, T., & Rodeheffer, C. (2018, July). Is This Person Real? Avatar Stylization and Its Influence on Human Perception in a Counseling Training Environment. In *International Conference on Virtual, Augmented and Mixed reality* (pp. 279-289). Springer, Cham.

Angelone, S., Soriguer, R. C., & Melendo, A. (2019). Modes of documentary films produced by the future generation of 'scientists-as-filmmakers.' *International Journal of Science Education, Part B*, *9*(4), 285-295.

Army Field Manual, FM 2-22.3 Human Intelligence Collector Operations, (2006), https://en.wikipedia.org/wiki/FM_2-22.3_Human_Intelligence_Collector_Operations and https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/fm2_22x3.pdf

Army Publishing Directorate (APD), (2022), Field Manuals, https://armypubs.army.mil/ProductMaps/PubForm/FM.aspx

Army Pubs, (2022), TRADOC Top 50 Web Views, https://armypubs.army.mil/pdf/apd_webviews.pdf

Blattgerste, J., Strenge, B., Renner, P., Pfeiffer, T., & Essig, K. (2017, June). Comparing conventional and augmented reality instructions for manual assembly tasks. In *Proceedings of the 10th international conference on pervasive technologies related to assistive environments* (pp. 75-82).

Bloom, Benjamin S. (March 1968). "Learning for Mastery." UCLA - CSEIP - Evaluation Comment. Vol. 1.

Branavan, S. R. K., Silver, D., & Barzilay, R. (2012). Learning to win by reading manuals in a Monte-Carlo framework. *Journal of Artificial Intelligence Research*, *43*, 661-704.

Brandon, S. E., Kleinman, S. M., & Arthur, J. C. (2019). A scientific perspective on the 2006 US Army Field Manual 2–22.3.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Bursztein, E., Bethard, S., Fabry, C., Mitchell, J. C., & Jurafsky, D. (2010, May). How good are humans at solving CAPTCHAs? A large-scale evaluation. In Security and Privacy (SP), 2010 IEEE Symposium on (pp. 399-413). IEEE.

Cambre, J., Reig, S., Kravitz, Q., & Kulkarni, C. (2020, July). "All Rise for the AI Director" Eliciting Possible Futures of Voice Technology through Story Completion. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (pp. 2051-2064).

Clearance.Jobs, (2021), "HUMINT vs. Counterintelligence- Diving Into the Disciplines," https://www.youtube.com/watch?v=-0O1HK5KoFY

Cohen, N. (2009), "Care to Write Army Doctrine? With ID, Log On", New York Times, https://www.nytimes.com/2009/08/14/business/14army.html

Datta, A., & Goswami, R. (2021). The Film Industry Leaps into Artificial Intelligence: Scope and Challenges by the Filmmakers. *Rising Threats in Expert Applications and Solutions* (pp. 665-670). Springer, Singapore.

Davis, E., & Marcus, G. (2015). The scope and limits of simulation in cognitive models. arXiv preprint arXiv:1506.04956.

Dobre, G. C. (2019, September). Using machine learning to generate engaging behaviours in immersive virtual environments. In *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)* (pp. 50-54). IEEE.

Engber, D. (2005) "What are Army Field Manuals?", Slate, https://slate.com/news-and-politics/2005/11/what-s-an-army-field-manual.html

Evin, I. (2021). Cine AI: Emulating Director Styles in Game Cutscenes.

Golla, R.G. (2022), Question Generator AI Repo., https://github.com/ramsrigouthamg/Questgen.ai

Graham, D. E. (2009). *A Long, Hard Fall from the Pedestal*. NATIONAL DEFENSE UNIV WASHINGTON DC INST FOR NATIONAL STRATEGIC STUDIES.

GlobalSecurity.org (2005), "Army Field Manuals," https://www.globalsecurity.org/military/library/policy/army/fm/

Hsieh, T. H., Tsai, Y. C., Kao, C. J., Chang, Y. M., & Lu, Y. W. (2013). A conceptual atomic force microscope using LEGO for nanoscience education. *International Journal of Automation and Smart Technology*, *4*(2), 113-121.

Johnson, J., Gupta, A., & Fei-Fei, L. (2018). Image generation from scene graphs. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1219-1228).

Kim, J. M., Hill Jr, R. W., Durlach, P. J., Lane, H. C., Forbell, E., Core, M., ... & Hart, J. (2009). BiLAT: A game-based environment for practicing negotiation in a cultural context. *International Journal of Artificial Intelligence in Education*, *19*(3), 289-308.

Kis, A., Tvaruška, P., & Tarcala, O. (2021, June). The Competing Hypotheses Analytical Model and Human Intelligence Single-Source Analysis. In *International conference KNOWLEDGE-BASED ORGANIZATION* (Vol. 27, No. 1, pp. 62-73).

Leonard, S. (2017), "Broken and Unreadable: Our Unbearable Aversion to Doctrine," https://mwi.usma.edu/broken-unreadable-unbearable-aversion-doctrine/

Magerko, B., Laird, J. E., Assanie, M., Kerfoot, A., & Stokes, D. (2004, July). AI characters and directors for interactive computer games. In *AAAI* (pp. 877-883).

Mansourifar, Hadi, and Weidong Shi. "One-shot gan generated fake face detection." *arXiv preprint arXiv:2003.12244* (2020).

Marcus, G., & Davis, E. (2020). Insights for AI from the human mind. Communications of the ACM, 64(1), 38-41.

Microsoft Read Aloud (2022), "Listen with reading Aloud," https://support.microsoft.com/en-us/office/listen-to-your-word-documents-5a2de7f3-1ef4-4795-b24e-64fc2731b001

Minor, J. (2021), "CapCut Review," PC Magazine, https://www.pcmag.com/reviews/capcut

MyHeritage.com (2022), DeepStory, https://www.myheritage.com/deepstory

Noever, D. (2022), Audio Book (3 minute sound file) production for FM2-22.3, pages 4-22, https://deeperbrain.com/film/WordTTS.ogg

Noever, D., Regian, W.(2022), Video Human Intelligence Collector Operations, https://deeperbrain.com/challenge/trainv6.mp4, based on 2006 Field Manual, https://armypubs.army.mil/ProductMaps/PubForm/Details.aspx?PUB_ID=82535

Noever, D., Regian, W.(2022), Cyber (CEMA) Field Manual FM 3-12, https://deeperbrain.com/challenge/cyberv3.mp4 , based on 2021 Field Manual, https://armypubs.army.mil/ProductMaps/PubForm/Details.aspx?PUB_ID=1022713

Noever, D., Regian, W.(2022), Chemical, Biological, Radiological, And Nuclear Operations, CBRNE Field Manual FM 3-11, https://deeperbrain.com/challenge/cbrnv1.mp4, based on 2019 Field Manual, https://armypubs.army.mil/ProductMaps/PubForm/Details.aspx?PUB_ID=1007035

Noever, D., Regian, W., Kalin, J., Ciolino, M. (2021) "AI The End of the World…of Work?", In Proceedings of Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Orlando, FL, 2 Dec-6 Dec 2021.

Noever, D., Regian, J.W. (2019) "Reinforcement Learning for Automated Textual Reasoning." In Proceedings of Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Orlando, FL, 2 Dec-6 Dec 2019.

Oleszkiewicz, S., Granhag, P. A., & Cancino Montecinos, S. (2014). The Scharff-technique: Eliciting intelligence from human sources. *Law and human behavior*, *38*(5), 478.

Open Media Foundation, (2022), Studio Rental Prices, http://openmediafoundation.org/studiorates

Pantserev, K. A. (2020). The malicious use of AI-based deepfake technology as the new threat to psychological security and political stability. In *Cyber defence in the age of AI, smart societies and augmented humanity* (pp. 37-55). Springer, Cham.

Parodi, A. (2013). *A Theoretically Driven Investigation of the Efficacy of an Immersive Interactive Avatar Rich Virtual Environment in Pre-deployment Nursing Knowledge and Teamwork Skills Training*. OLD DOMINION UNIV RESEARCH FOUNDATION NORFOLK VA.

Rackaway, C. (2012). Video killed the textbook star?: Use of multimedia supplements to enhance student learning. *Journal of Political Science Education*, *8*(2), 189-200.

Regian, J. W., Dennis, M. M., & Shute, V. J. (1989). *INFLITE: An Intelligent Instrument Flight Trainer with Computer-Generated Speech*. AIR FORCE HUMAN RESOURCES LAB BROOKS AFB TX.

Regian, W., & Pitts, G. (1992, September). A Fuzzy Logic-Based Intelligent Tutoring System (ITS). In *Proceedings of the IFIP 12th World Computer Congress on Education and Society-Information Processing'92-Volume 2-Volume 2* (pp. 66-72).

Regian, J. W., & Shute, V.J. (2013). *Cognitive approaches to automated instruction*. Routledge.

Regian, J. W. & Shute, V. J. (1992). Cognitive approaches to automated instruction. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rubens, N., Kaplan, D., & Okamoto, T. (2012, September). E-Learning 3.0: anyone, anywhere, anytime, and AI. In *International conference on web-based learning* (pp. 171-180). Springer, Berlin, Heidelberg.

Sheth, V., Bokshi, M., Pruett, E., Drotar, C., Wan, Y., Fu, S., & Namuduri, K. (2013). A testbed for multi-domain communication networks using LEGO Mindstorms. In *AIAA Infotech@ Aerospace (I@ A) Conference* (p. 5056).

Steward, S. A. (2004). A Rhetoric of Technology: The Discourse in US Army Manuals and Handbooks.

Stockwell, B. R., Stockwell, M. S., Cennamo, M., & Jiang, E. (2015). Blended learning improves science education. Cell, 162(5), 933-936.

SwiftRead (2022), "How long does it take to read…?" https://swiftread.com/reading-time/384-pages

Talbot, C. (2013, May). Creating an artificially intelligent director (aid) for theatre and virtual environments. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems* (pp. 1457-1458).

Tang, Y., Ding, D., Rao, Y., Zheng, Y., Zhang, D., Zhao, L., ... & Zhou, J. (2019). Coin: A large-scale dataset for comprehensive instructional video analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1207-1216).

Temiz, Y. (2020). The LEGO microscope: A valuable lab tool began as a DIY project-[Hands on]. *IEEE Spectrum*, *57*(5), 16-18.

Vos, B. E., Blesa, E. B., & Betz, T. (2021). Designing a high-resolution, LEGO-based microscope for an educational setting. *The Biophysicist*, *2*(3), 29-40.

Yannakakis, G. N. (2012, May). Game AI revisited. In *Proceedings of the 9th conference on Computing Frontiers* (pp. 285-292).

YouTube, (2022), search for "human intelligence collection mos" and top 12 returned video links, see https://www.youtube.com/results?search_query=human+intelligence+collector+mos

Wahlquist, J. A. (2009). *Enhancing interrogation: Advancing a new agenda*. ARMY WAR COLL CARLISLE BARRACKS PA.

Williams, M., & Smith, R. (2002). The Development of an Intelligence and Electronic Warfare FOM to Bridge Constructive, Virtual, and Live Simulations. In the *2002 Spring Simulation Interoperability Workshop*.

Winograd, A (2022) "How much does it cost to hire an actor?" Veedme Knowledge Base, Video Production https://www.veed.me/how-much-does-it-cost-to-hire-an-actor/