ChatGPT

AI Models for Detecting Team Behaviors in Virtual Meetings

Introduction

Analyzing team interactions in virtual meetings (e.g. Zoom) has become a rich area of research, with AI models developed to automatically detect various team behaviors and processes. These behaviors span **communication strategies** (what is said and how), **nonverbal cues** (visual gestures and actions), **speaking patterns** (turn-taking and talk time), and **emotional expressions**. Such behaviors are often correlated with team success – for instance, high engagement and positive communication often lead to better outcomes people.csail.mit.edu people.csail.mit.edu . Recent work has even trained multimodal AI systems to classify overall **teamwork engagement**, achieving over *91% accuracy* in distinguishing "productive" vs. "unproductive" team sessions from combined audio, video, and text streams of online meetings people.csail.mit.edu . Below, we review the most-cited models in each behavior category, highlighting their approach, performance, and popularity, and noting gaps where certain behaviors lack robust models.

Communication Strategies and Conversational Acts

Detecting Proposals, Decisions and Critiques: One key communication process is how teams propose and decide on ideas. Researchers from the AMI project built classifiers to detect **decision-making segments** in meeting transcripts. For example, Hsueh *et al.* developed the **"AMI DecisionDetector"**, which uses lexical cues (words), dialogue-act tags, and prosody to mark segments where decisions are made idiap.ch idiap.ch . Their model, trained on 50 multi-party meetings, achieved about **68% F1-score** in identifying decision-related segments on unseen meetings idiap.ch . Lexical features were most predictive, but adding dialogue-act context (e.g. whether an utterance is an *Inform*, *Suggest*, etc.) improved precision by filtering out non-decision content idiap.ch idiap.ch . This work is highly cited in meeting summarization research for its contribution to **automatic meeting understanding**.

Agreement and Disagreement Detection: Another well-studied strategy is the expression of agreement or criticism. Hillard, Ostendorf & Shriberg's influential model (cited ~300 times) classifies **agreement vs. disagreement utterances** in multiparty meetings aclanthology.org . By combining words (NLP) with **prosodic cues** (tone, intonation), their system could **recover ~80% of agreement/disagreement utterances** (at ~3% false alarm rate) even on noisy automatic speech transcripts aclanthology.org . For example, an utterance like *"This doesn't answer the question."* would be flagged as a negative disagreement aclanthology.org . They also identified trivial backchannel words (e.g. "uh-huh", "yeah") as a separate class since these short acknowledgments often **encourage the speaker to continue** rather than signal true agreement aclanthology.org . This early work demonstrated that detecting subtle **speech acts** (like disagreeing or encouraging) is feasible with machine learning. It remains a reference point, inspiring later deep-learning models for dialog act classification.

Role and Discussion Pattern Analysis: Beyond individual utterances, AI has been used to infer each member's **communicative role** or style in the team. A notable example is Dowell *et al.*'s **Group Communication Analysis (GCA)** framework link.springer.com . Using computational linguistics on large online team discussion datasets (N≈2,400 participants), they clustered communication behaviors into emergent roles like "Driver" (active leader), "Chatterer" (high volume but low cohesion), "Lurker" (minimal contributor), etc. link.springer.com link.springer.com . These roles were shown to predict performance – e.g. teams with more **"Influential Actor"** roles (fewer contributions but high relevance and attentiveness) performed as well as those dominated by Drivers link.springer.com link.springer.com . Another line of work focuses on **social roles** in meetings: for instance, research by Valente *et al.* recognized four **socio-emotional roles** – **Protagonist** (dominant speaker), **Supporter**, **Neutral**, **Gatekeeper** – using audio features and turn-taking patterns isca-archive.org isca-archive.org . Their model (an HMM-based classifier on the AMI corpus) reached **68% accuracy** at labeling these roles isca-archive.org . Notably, **Supporter** and **Gatekeeper** behaviors (which correspond to encouraging others and mediating turn-taking) were hardest to detect due to infrequent examples isca-archive.org . Overall, while content-based **proposal/decision detection** and **agreement classification** are well-covered by popular models, identifying nuanced facilitative acts (like explicitly *inviting input* or *encouraging quiet members*) remains a gap – these often get lumped into broader categories or require manual coding in studies.

Nonverbal Behaviors and Visual Cues

In virtual meetings, visual team behaviors are limited to what webcams capture – primarily faces, head movement, and some hand gestures. Despite this, researchers have made progress in detecting **gestures and visual engagement cues** with AI.

Head Nods and Gestures: Head nods are important signals of engagement and agreement. Computer vision models (often HMM or deep-learning based) can detect head nods and shakes in real time. For example, Artiran *et al.* (2021) proposed an HMM-based detector for nods using head motion data, achieving high agreement with human coders on **conversational engagement** scores eurasip.org eurasip.org . Real-time vision systems dating back to Morency (2005) also showed accurate nod/shake detection (>90% precision) using facial feature tracking sciencedirect.com . These models are useful for gauging **active listening** – e.g. frequent nodding by listeners often correlates with higher participation and understanding. **Hand gestures** are harder to capture in webcam views, but proxies exist. Hung *et al.* (2007) introduced a clever approach to estimate **body movement levels** from compressed video streams idiap.ch idiap.ch . They measured each participant's motion via changes in video encoding (motion vectors and residuals) and used it to predict dominance idiap.ch . Although audio alone was a stronger predictor of who was most active, adding these **visual motion features** modestly improved accuracy in identifying the most dominant person in the meeting idiap.ch idiap.ch . This suggests that even limited **gestural activity** data can inform models about behaviors like who is **leading** (often moving more, gesturing while talking) vs. who is passive.

Screen Sharing and Gaze: Direct detection of **screen sharing** events has not been a focus of academic vision research – likely because these events are explicit (the video feed changes or an icon appears). In practice, meeting platforms log screen-share usage. Thus, while an AI **model isn't needed to detect when screen-sharing happens**, the *behavioral significance* can be analyzed (e.g. teams that frequently share their screen might be more task-focused). **Gaze and eye contact**, important in in-person settings, are challenging to infer over Zoom (everyone's gaze appears off due to cameras). Few AI models tackle eye-contact in Brady Bunch–style video grids, making this a gap in current capabilities. Researchers instead rely on head orientation or attention trackers to estimate who is looking at whom, but robust solutions are not yet common for gallery-view video.

Nonverbal Synchrony: An exciting development in team research is measuring **behavioral synchrony** – how in-sync team members' physical motions are. Highly cited works by Pentland's group and others have shown that synchronous nodding, posture shifts, or rhythm in speaking correlates with rapport and performance. A recent open-access study by Manabe *et al.* (2024) used instrumented chairs (the **SenseChair**) to quantify body-motion synchrony in group discussions nature.com. They found a **positive correlation between synchrony and idea generation**: groups where more pairs of teammates moved in rhythm produced significantly more brainstorming ideas nature.com. In other words, when **team members unconsciously mirror each other's nods or gestures**, the group tends to be more creative and productive. This aligns with prior findings that interactional synchrony fosters cooperation. For virtual meetings, AI can approximate synchrony by analyzing concurrent nodding or simultaneous facial expressions across video feeds. Some pilot systems track multi-person facial landmarks to compute a synchrony score (though these are experimental). Overall, **visual cues like nods, gestures, and synchrony** are being detected with reasonable accuracy and proven relevant – but certain in-person behaviors (full-body language, direct eye contact) remain **partially observable or untapped by AI** in virtual contexts.

Speaking Duration and Turn-Taking Patterns

Conversation dynamics – who talks when and for how long – are strong indicators of team functioning. Several classic AI models focus on these audio features, as they are readily extracted from meeting recordings.

Speaking Time & Dominance: One fundamental metric is each member's total **speaking duration**. A well-known study by Hung *et al.* (ACM MM 2007) showed that simply using total speaking time can predict the **most dominant person** in a meeting with about **82% accuracy** idiap.ch . In their experiments on 4-person meetings, the participant with the longest talking time (and highest voice energy) was correctly identified as the "dominant" in most cases idiap.ch . This result, widely cited in social signal processing, confirms that imbalance in speaking time is a telltale sign of dominance or leadership taking hold. Conversely, successful teams often exhibit *balanced* participation – an insight echoed in "collective intelligence" studies (e.g. Woolley et al., 2010) and supported by AI analytics. In fact, Pentland's **Sociometric Badge** research found that teams with more equal turn-taking had higher productivity and creativity (though that work used wearable sensors rather than Zoom audio). Modern AI can compute these features from audio: voice activity detection and diarization algorithms segment who spoke when, enabling metrics like talk fraction, interruption counts, and overlap time. Such models are standard and very accurate in clean audio conditions (speaker diarization on known meeting corpora can exceed 90% accuracy in segmenting speakers).

Turn-Taking and Coordination: Beyond amounts of speech, **patterns of exchange** matter. A 2021 PLOS ONE study by Tomprou *et al.* examined *turn-taking synchrony* in remote teams journals.plos.org journals.plos.org . Surprisingly, they found that **teams communicating with audio-only (no video)** achieved *greater vocal synchrony* (more fluid, evenly spaced turn exchanges) and consequently **higher collective intelligence scores** than teams with video journals.plos.org . When video was on, teams tended to rely less on vocal cues and had more uneven turn-taking. This suggests that **coordinating turn exchanges** (not talking over each other, responding in rhythm) is crucial for group effectiveness, and AI can measure this. Their model computed metrics like turn-taking lag, simultaneous speech incidents, and prosodic alignment, linking those to the team's problem-solving scores. Another example, Gervits *et al.* (2016), treated **team communication as a collaborative process** and showed that concise, well-coordinated turn-taking (few long monologues, more interactive back-and-forth) was associated with better team performance link.springer.com link.springer.com . Machine learning models can classify certain patterns – e.g. detecting if a conversation is **one-sided** or **interactive** – using features such as the number of speaker switches, average turn length, and silence gaps. Such classifiers can effectively flag dysfunctional meetings (e.g. one person monologuing) versus highly interactive ones. Overall, **audio-based models** for speaker diarization, dominance ranking, and interaction smoothness are well-established and widely used (some integrated into meeting tools). A gap remains in higher-level inference – for instance, automatically recognizing **interruptions vs. smooth hand-offs** or detecting when a facilitator is **inviting others to speak**. These often require combining timing with linguistic cues.

Emotional Expressions and Sentiment

Team success can also hinge on the emotional tone of meetings. AI models have been developed to detect **facial expressions, vocal emotion, and overall sentiment**, enabling analysis of how positive or negative the team interaction is.

Facial Emotion Recognition: Modern deep learning has made facial emotion detection (e.g. smiles, frowns) quite reliable. Models like OpenFace or EmoNet (2010s) can classify basic expressions (happiness, frustration, etc.) from webcam video with 90%+ accuracy under good conditions. In team contexts, **smiling and laughter** are especially relevant as signs of positivity and cohesion. A recent work by Bohy *et al.* (2024) highlights advances in **smile and laugh detection** arxiv.org . They created a multimodal CNN that treats smiles and laughter as distinct classes, using both video (face movements) and audio (chuckles) arxiv.org arxiv.org . The fused model outperformed vision-only ones, especially at distinguishing subtle smile-versus-laugh intensity arxiv.org . This reflects the trend in emotion AI: **multimodal fusion** is best, since laughter has both audible and visible components. Previous studies (e.g. Kantharaju 2018) similarly achieved high accuracy (>85% F1) in detecting **laughter episodes** in meetings by combining audio (spectral features) with visual cues arxiv.org . These models let researchers quantify team humor – e.g. measuring how often the group laughs together, which can indicate rapport. They also detect **signs of frustration or anger** via facial cues like frowns or head shakes. For instance, an "anger" expression detector (based on FACS action units) can flag if a participant consistently looks frustrated. However, privacy and video quality can limit usage of face-emotion models in real meetings.

Vocal Tone and Sentiment: In parallel, **speech emotion recognition (SER)** has matured. AI can analyze tone of voice to gauge emotions such as excitement, boredom, or irritation. For example, Mel-frequency cepstral coefficients (MFCCs) fed into an RNN or transformer can classify an utterance's emotion with around 70–80% accuracy in acted datasets cacm.acm.org . In meetings, a practical application is **sentiment analysis** on transcripts: using NLP to detect positive vs negative language. If team members use many positive words ("great idea!", "I agree") versus negative or anxious language ("I'm not sure", "this is bad"), it sets the emotional climate. One study on group **emotion and cohesion** (Lim *et al.*, 2024) found that meetings labeled with **positive group emotion** (everyone feeling upbeat) almost always had high cohesion ratings, whereas **negative-emotion meetings** corresponded to low cohesion mdpi.com . This underscores that **a positive emotional atmosphere correlates with better teamwork**. AI models can support this by detecting collective sentiment: e.g. aggregating each speaker's predicted sentiment per turn to compute a positivity ratio for the meeting. High-quality pre-trained language models (like BERT-based classifiers) now achieve >90% accuracy on sentiment of general text, though for **meeting dialogue** the accuracy may be slightly lower due to informal language. Still, these are widely used (and cited) tools in industry meeting analytics (e.g. Microsoft Teams' sentiment insights).

Observations and Gaps: Most emotional behavior detection models perform well on discrete labels (happy/sad or positive/negative). They have been benchmarked on public datasets (the IEMOCAP speech corpus, facial expression databases, etc.) with robust results. In meetings research, they are popular for assessing **team morale and engagement**. For instance, monotonic voice and lack of smiling might signal disengagement, while frequent smiling and energetic tone signal **enthusiasm** – models can capture these signals and have been validated against human ratings of engagement eurasip.org eurasip.org . A gap in this category is detecting more nuanced affects like **sarcasm or humor context** (beyond just laughter). Also, **frustration** can manifest subtly (sighs, eye-rolls) that current classifiers might miss or mislabel. High-quality preprints are exploring these (e.g. multi-modal sarcasm detection in video meetings), but they aren't yet standard. Additionally, ensuring emotion models generalize across cultures and individuals (since expressions vary) is an ongoing challenge.

Conclusion and Model Recommendations

Across these categories, the **most successful models** tend to be those that leverage **multiple modalities** and have been evaluated on real meeting data. For communication content, **NLP classifiers** for dialog acts (decisions, agreements) are highly cited and show strong performance (often 70–80% accuracy) aclanthology.org . For nonverbal cues, **vision-based detectors** (nod/shake, motion energy) and synchrony measures have proven useful, though some behaviors (like full body language or screen gaze) remain hard to capture in Zoom. Audio-based models for **speaking patterns** are very mature, with simple features like talk time yielding reliable indicators of dominance and balanced involvement idiap.ch . Emotion recognition, through both **facial analysis and vocal sentiment**, has been benchmarked extensively; these models can automatically gauge team positivity/frustration with reasonable accuracy, which correlates with team cohesion mdpi.com .

In terms of popularity, the models emerging from large projects (AMI Meeting Corpus, ICSI, MIT's Teamwork project) are among the most cited. For example, the agreement/disagreement detector by Hillard *et al.* and the dominance detection by Hung *et al.* are cornerstone references in computational meeting analysis. **Performance-wise**, integrated systems (combining signals) report the best results – e.g. the 91% engagement classifier combining audio, video, text people.csail.mit.edu . We recommend an **ensemble approach**: using a suite of specialized models – one for conversational acts (to catch idea proposals or critiques), one for visual engagement (nods, attention), one for turn-taking metrics, and one for emotional tone. Each model can contribute a piece to the overall picture of team behavior. Notably, some behaviors lack dedicated models (e.g. *explicitly encouraging others to participate* is not directly classified by any off-the-shelf model). These gaps present opportunities to train new AI detectors, perhaps using labeled data of such behaviors from team meeting recordings.

By deploying top-performing models in each category – for example, an NLP dialog-act model for proposals idiap.ch , a vision model for nodding eurasip.org , an audio-turn-taking analyzer journals.plos.org , and an emotion recognizer arxiv.org – researchers can automatically quantify many key team behaviors. This enables examining how those behaviors correlate with success (e.g. does high nod synchrony and equal talk time predict higher innovation output?). In summary, AI research provides a growing toolkit to observe **what effective teams do**, even in virtual settings. The most cited and validated models already detect many crucial behaviors with good accuracy, while a few nuanced behaviors still await robust modeling – an exciting frontier for future work in **automated team analytics**.

Sources: The information above is synthesized from prominent research papers and surveys on automated meeting analysis, including the IEEE/ACM Transactions on multimedia meetings, ACL meeting understanding workshops, and recent open-access studies in applied sciences and social computing idiap.ch aclanthology.org idiap.ch mdpi.com , among others. These works benchmarked model performance and often correlated their detections with team outcomes, providing a foundation for selecting an ensemble of AI models for comprehensive team behavior measurement.

Citations

AI-Based Automatic Detection of Online Teamwork Engagement in Higher...
https://people.csail.mit.edu/dkao/pdf/AI-Based_Automatic_Detection_of_Online_Teamwork_Engagement_in_Higher_Education.pdf

AI-Based Automatic Detection of Online Teamwork Engagement in Higher...
https://people.csail.mit.edu/dkao/pdf/AI-Based_Automatic_Detection_of_Online_Teamwork_Engagement_in_Higher_Education.pdf

AI-Based Automatic Detection of Online Teamwork Engagement in Higher...
https://people.csail.mit.edu/dkao/pdf/AI-Based_Automatic_Detection_of_Online_Teamwork_Engagement_in_Higher_Education.pdf

https://www.idiap.ch/webarchives/sites/publications.amiproject.org/lnai-hsueh-final-errata.pdf

https://www.idiap.ch/webarchives/sites/publications.amiproject.org/lnai-hsueh-final-errata.pdf

https://www.idiap.ch/webarchives/sites/publications.amiproject.org/lnai-hsueh-final-errata.pdf

https://www.idiap.ch/webarchives/sites/publications.amiproject.org/lnai-hsueh-final-errata.pdf

https://aclanthology.org/N03-2012.pdf

https://aclanthology.org/N03-2012.pdf

https://aclanthology.org/N03-2012.pdf

https://aclanthology.org/N03-2012.pdf

Group communication analysis: A computational linguistics approach for ...
https://link.springer.com/article/10.3758/s13428-018-1102-z?error=cookies_not_supported&code=0d0cd24e-43ea-43cd-8890-822898c0c7a2

Group communication analysis: A computational linguistics approach for ...
https://link.springer.com/article/10.3758/s13428-018-1102-z?error=cookies_not_supported&code=0d0cd24e-43ea-43cd-8890-822898c0c7a2

Group communication analysis: A computational linguistics approach for ...
https://link.springer.com/article/10.3758/s13428-018-1102-z?error=cookies_not_supported&code=0d0cd24e-43ea-43cd-8890-822898c0c7a2

Group communication analysis: A computational linguistics approach for ...
https://link.springer.com/article/10.3758/s13428-018-1102-z?error=cookies_not_supported&code=0d0cd24e-43ea-43cd-8890-822898c0c7a2

https://www.isca-archive.org/interspeech_2011/valente11b_interspeech.pdf

https://www.isca-archive.org/interspeech_2011/valente11b_interspeech.pdf

https://www.isca-archive.org/interspeech_2011/valente11b_interspeech.pdf

https://eurasip.org/Proceedings/Eusipco/Eusipco2021/pdfs/0001301.pdf https://eurasip.org/Proceedings/Eusipco/Eusipco2021/pdfs/0001301.pdf

A real-time head nod and shake detector using HMMs - ScienceDirect
https://www.sciencedirect.com/science/article/abs/pii/S0957417403000885

https://www.idiap.ch/~gatica/publications/HungJayagopiYeoFriedlandBaOdobezRamchandra
acmmm07.pdf

https://www.idiap.ch/~gatica/publications/HungJayagopiYeoFriedlandBaOdobezRamchandra
acmmm07.pdf

https://www.idiap.ch/~gatica/publications/HungJayagopiYeoFriedlandBaOdobezRamchandra
acmmm07.pdf

https://www.idiap.ch/~gatica/publications/HungJayagopiYeoFriedlandBaOdobezRamchandra
acmmm07.pdf

https://www.idiap.ch/~gatica/publications/HungJayagopiYeoFriedlandBaOdobezRamchandra
acmmm07.pdf

The association between synchrony and intellectual productivity in a grou...
https://www.nature.com/articles/s41599-023-02566-1?
error=cookies_not_supported&code=881ffd8f-606a-4924-b256-330e90236a2a

The association between synchrony and intellectual productivity in a grou...
https://www.nature.com/articles/s41599-023-02566-1?
error=cookies_not_supported&code=881ffd8f-606a-4924-b256-330e90236a2a

https://www.idiap.ch/~gatica/publications/HungJayagopiYeoFriedlandBaOdobezRamchandra
acmmm07.pdf

Speaking out of turn: How video conferencing reduces vocal synchrony a...
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0247655

Speaking out of turn: How video conferencing reduces vocal synchrony a...

Speaking out of turn: How video conferencing reduces vocal synchrony a...
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0247655

Speaking out of turn: How video conferencing reduces vocal synchrony a...
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0247655

Group communication analysis: A computational linguistics approach for ...
https://link.springer.com/article/10.3758/s13428-018-1102-z?error=cookies_not_supported&code=0d0cd24e-43ea-43cd-8890-822898c0c7a2

Group communication analysis: A computational linguistics approach for ...
https://link.springer.com/article/10.3758/s13428-018-1102-z?error=cookies_not_supported&code=0d0cd24e-43ea-43cd-8890-822898c0c7a2

A New Perspective on Smiling and Laughter Detection: Intensity Levels M...
https://arxiv.org/html/2403.02112v1

A New Perspective on Smiling and Laughter Detection: Intensity Levels M...
https://arxiv.org/html/2403.02112v1

A New Perspective on Smiling and Laughter Detection: Intensity Levels M...
https://arxiv.org/html/2403.02112v1

A New Perspective on Smiling and Laughter Detection: Intensity Levels M...
https://arxiv.org/html/2403.02112v1

A New Perspective on Smiling and Laughter Detection: Intensity Levels M...
https://arxiv.org/html/2403.02112v1

Speech Emotion Recognition - Communications of the ACM
https://cacm.acm.org/research/speech-emotion-recognition/

GCE: An Audio-Visual Dataset for Group Cohesion and Emotion Analysis
https://www.mdpi.com/2076-3417/14/15/6742

https://eurasip.org/Proceedings/Eusipco/Eusipco2021/pdfs/0001301.pdf

A New Perspective on Smiling and Laughter Detection: Intensity Levels M...
https://arxiv.org/html/2403.02112v1

All Sources

people.csail.mit idiap aclanthology link.springer isca-archive

eurasip sciencedirect nature journals.plos arxiv cacm.acm

mdpi