The Future of Emotion Recognition in Machine Learning

By Martin Anderson

First published ***October 21st, 2020*** *at:*

https://www.iflexion.com/blog/emotion-recognition-software

Web-archived version

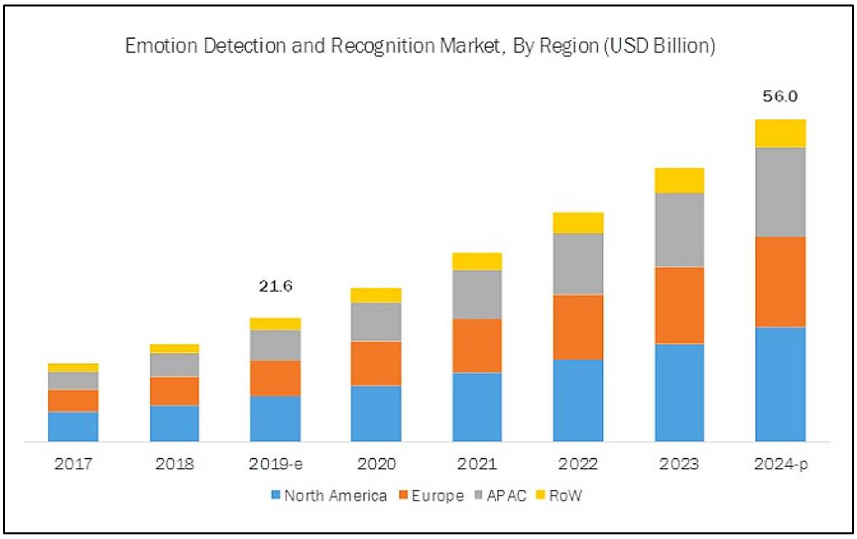
Understanding the meaning of facial expressions is an essential human survival tool, but one that generates scientific controversy. Some contend that we recognize anger most easily[1], since it may represent an existential threat; others, that we recognize happiness most easily and anger most slowly[2], since the march of civilization has altered our priorities over time.

The lexicon of facial expression is rich in ambiguities, as facial cues may be insincere, misinterpreted, or in some way at a tangent to their intent. Experts in the understanding of facial expression even remain split on whether the Mona Lisa's enigmatic smile is sincere[3] or forced[4] (apparently a matter of context).

The ability to decipher the true intent and emotional response of a person from their facial expressions, notwithstanding their attempts to mask or deceive what they feel, is an evolutionary advantage of great interest to a range of sectors, from physicians through to marketers and political analysts. Unsurprisingly, there's a lot of money in it.

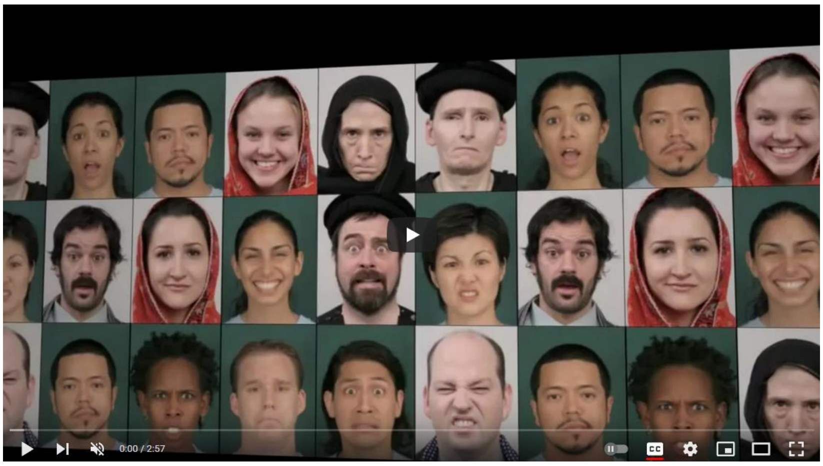
Market Growth in Facial Expression Recognition Technologies

Technologies based on Facial Expression Recognition (FER, also known as Affect Recognition) form a significant part of the emotion recognition market, estimated to reach a value of $56 billion by 2024[5].



Source: https://www.marketsandmarkets.com/Market-Reports/emotion-detection-recognition-market-23376176.html

The advent of industrialized machine learning techniques has combined with high availability of sample imagery and a huge commercial investment in facial recognition software to generate new FER technologies that may augur scientific, governmental and commercial access to our innermost thoughts.



Despite that possibility, these new capabilities raise some difficult questions. Can computer vision software account for the differences in expression across race, age and gender? Is FER a solution in search of a problem? Are growing concerns around privacy and governance a significant obstacle to such technologies?

And does the dominant theory of 'expression recognition' itself have enough scientific credibility to be perpetuated into machine learning systems?

Uses for Expression Recognition

DeepFake Detection

Academic research into identifying DeepFake video manipulations ramped up in the 12 months prior to the 2020 U.S. election, in response to widely-voiced concerns[6,7,8] around deceptive video content with a political agenda.

In 2019 the Computer Vision Foundation partnered with UC Berkley, Google and DARPA to produce a system claimed to identify DeepFake manipulations via the analysis of expressions in the targeted subjects[9] – the only one, among several novel approaches[10], that relies specifically on facial expressions.

From left to right: the target subject, impersonator and 'deepfaked' impersonator. Source: https://openaccess.thecvf.com/content_CVPRW_2019/papers/Media%20Forensics/Agarwal_Protecting_World_Leaders_Against_Deep_Fakes_CVPRW_2019_paper.pdf

Because it is more difficult to obtain a high volume of images of 'rarer' expressions for a machine learning data-set, these will inevitably be under-represented at the training stage, and will therefore be more detectable in comparison to 'stock' expressions in a DeepFake video, which were trained on a higher number of expression instances.

Medical Research Into Autism

Machine learning systems for facial emotion recognition are particularly suited for the study of Autism Spectrum Disorder (ASD), where sufferers have developmental and long-term difficulties in evaluating facial emotions[11].

One 2018 paper[12] from India leverages FER by processing publicly available social media images through a workflow involving TensorFlow, NumPy, OpenCV and Dlib, to generate labeled images that can be shown under laboratory conditions to ASD sufferers in order to help them improve their expression recognition skills.

This project uses Haar cascade object detection to identify and label expressions for use in autism spectrum disorder research and therapy. Source: http://www.ijitee.org/wp-content/uploads/papers/v9i7/F3772049620.pdf

Stanford University's Autism Glass Project[13] uses Google's face-worn computing system to aid autism-affected subjects in understanding appropriate social cues by appending emoticons to those expressions it is able to recognize.

The Autism Glass Project: An emoticon appears to the Glass-wearing subject when the system identifies a facial expression. Source: https://www.youtube.com/watch?v=_kzfuXy1yMI

Another project[14] has used machine learning to develop an app to screen children for autism by running the subject's facial reactions to a movie through a behavioral coding algorithm in order to identify the nature of their responses (in comparison to a person not afflicted with ASD). The project utilizes Amazon Web Services (AWS), PyTorch and TensorFlow.

Video: *Researchers Are Using Machine Learning to Screen for Autism in Children* - https://www.youtube.com/watch?v=YQpTlnWYAqE

Since autism is such an apt candidate for expression recognition technologies, there are more applications of machine learning for this cause[15] than we are able to list here.

Automotive Safety and Research Systems

Due to increased controversy over data mining and privacy concerns[16] over the last five years, the diffusion and extent of always-on driver monitoring systems has hit a number of roadblocks, relative to press exposure and anticipation[17] around such technologies.

Nonetheless, expression recognition is included as part of a number of available in-car systems trained by machine learning. For instance, in addition to its ability to understand if a driver is not looking at the road, not wearing a mask or is making a hands-on phone call, Affectiva's in-cabin sensing suite[18] includes an expression-recognition component that's capable of determining if the driver is falling asleep.

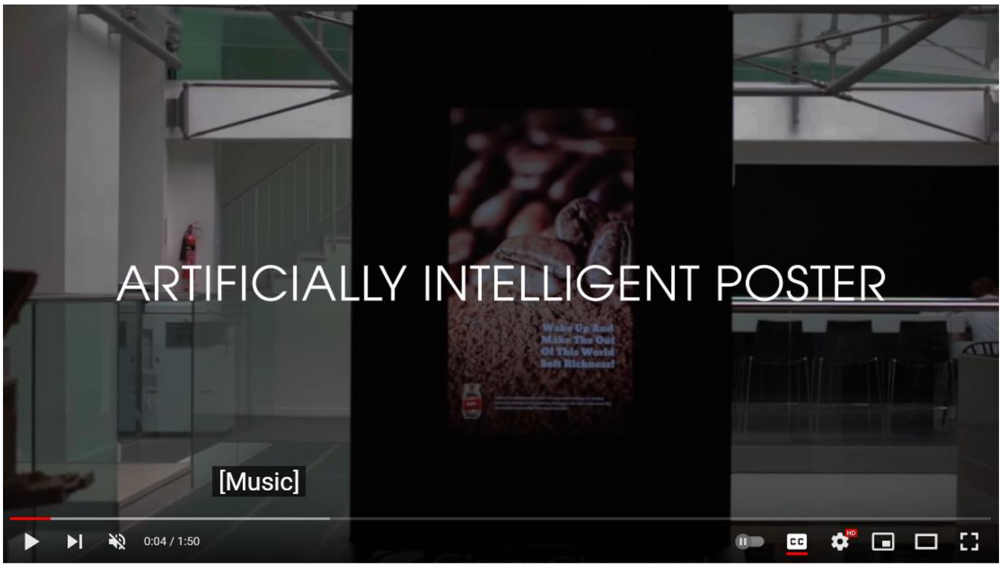
Source: https://www.youtube.com/watch?v=oSBmMT2dWFo

This test for 'liveness' or alertness can be configured to trigger a number of warnings or other safety-related actions.

Market Research

Since monitoring user data by stealth has become such a political football in recent years, facial expression recognition technologies are on their strongest ground where the subjects are willing accomplices, such as in focus groups and other forms of beta-testing for product marketing.

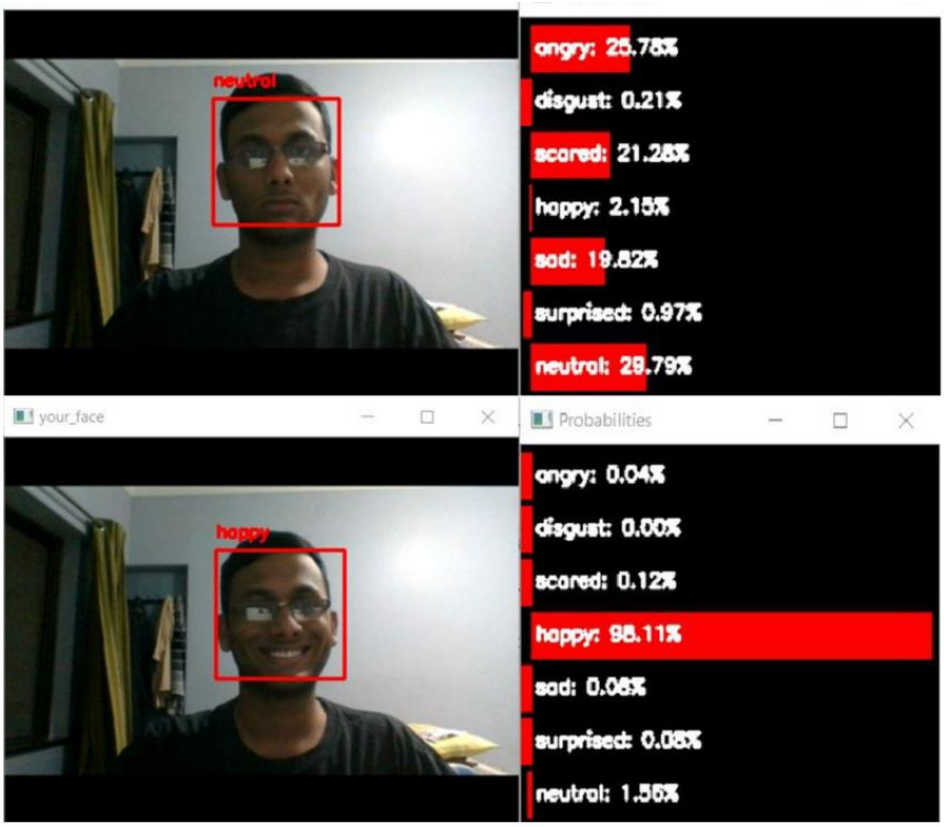
Outside of such controlled environments, there are possibilities for novel use of FER in a marketing context, such as M&C Saatchi's 2015 experiment[19] with an AI-powered poster that changed the advertising content dependent on the expressions that it could read from passers-by.



Video: *Artificially Intelligent Poster*- https://www.youtube.com/watch?v=JfpuqfqC-ts

Recruitment

The rise of teleconferencing-based interaction since the advent of COVID-19 solves two major obstacles to using FER in a recruitment interview context: the negative effects[20] of explicitly training a camera on someone you are talking to (solved by the user's webcam); and the more recent challenge of the subject's face being occluded by a mask.



A real-time FER classification system in action. Source: http://j-asc.com/gallery/106-may-2674.pdf

Though FER is being increasingly implemented in a recruitment context[21],[22], it comes with a number of caveats: depending on the system, it can introduce racial bias[23] that can expose employers to legal repercussion (a sub-set of wider problems around ethnicity in facial recognition[24]); it will probably work best without user consent[25] – usually illegal in the EU and in many other jurisdictions; and it seems set for further regulation[26,27], potentially compromising any current effort of investment.

Besides these considerations, an FER-based recruitment system can eventually be 'gamed': in South Korea, where FER is increasingly used in the hiring process, and where a quarter of the top 131 companies either use or plan to use AI in recruitment[28], candidates are now schooling themselves[29] in anti-FER techniques.

Pay-Per-Laugh

Faced with rising ticket prices and declining audiences In 2014, a Spanish theatre experimented with charging audiences 30 cents a laugh for its comedy shows[30], installing emotion recognition cameras in front of each seat, and fixing a ceiling of 24 euros per customer for the more successful events.

Video: 'PAY PER LAUGH -The CYRANOS McCANN Woldgroup Europe Barcelona Cannes Lions 2014 Winner' - https://www.youtube.com/watch?v=V0FowbxEe3w

Despite the viral spread of the project, 'metered' comedy shows have not stormed the market since, though some have speculated that the principle could be applied for other types of output, such as horror.

Facial Expression Implementations and Products

Though cloud-based services are usually unsuitable for time-critical deployments (such as in-car safety systems), they are useful for training a more responsive and slimmed-down algorithm, or else in evaluating data in research projects where latency is not a factor.

The commercial SkyBiometry API, which provides a range of facial detection and analysis features, can also individuate[31] anger, disgust, neutral mood, fear, happiness, surprise and sadness. Microsoft Azure's Emotion API[32] can also return emotion recognition estimates along with the usual array of feature requests.

However, the two volume commodity providers emerging from this circumspect market are Google Cloud Vision[33] and Amazon Rekognition[34], both of which provide facial sentiment detection facilities as a component of their more broadly successful facial recognition APIs.

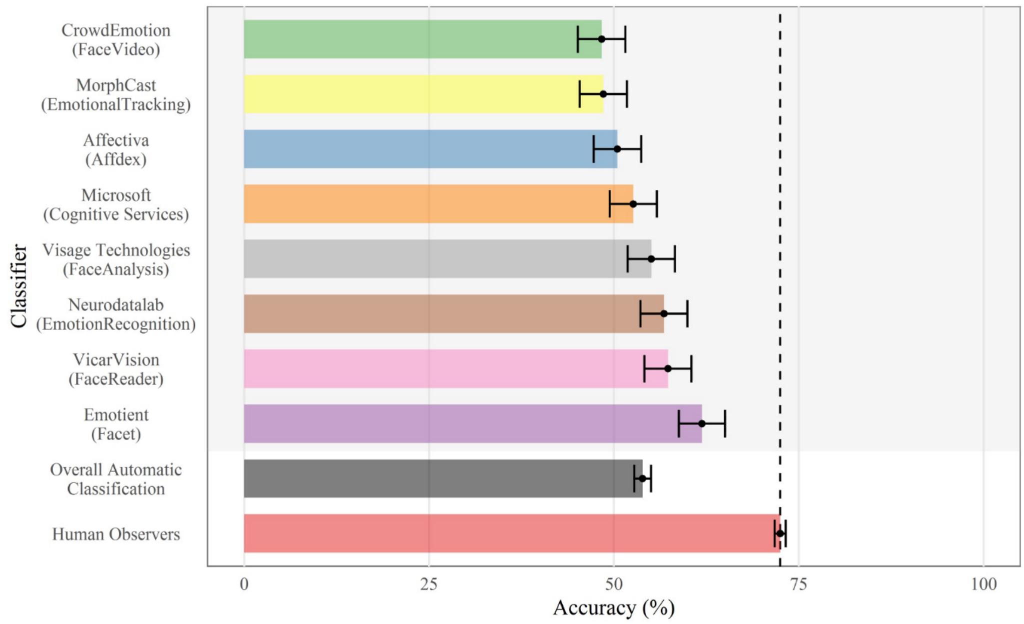
Though their comparative offerings are not identical in terms of properties or methods of quantifying success, one 2019 comparison[35] of the services found that Google was less likely to commit to identifying an emotion at all, whereas Rekognition is willing to commit to an emotion at levels as low as 5% in order to return a result of some kind.

Emotient

Apple was patenting emotion recognition mechanisms as far back as 2012[36]. However, besides an expression-based emoji system[37], the tech giant has delivered little working FER functionality in the wake of its acquisition of artificial intelligence startup Emotient in 2016[38].

However, a more recent 2019 patent[39] has led speculation[40] that Apple's Siri assistant may soon use an FER methodology to identify user emotions via facial expressions, though in combination with voice tonality analysis. If true, it's unclear as of yet what Siri would do with this information.

One 2020 study[41] compared eight commercially available FER systems against human responses to facial expressions, with Apple's Emotient system scoring highest.



Source: https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0231968&type=printable

The Challenge of Recognizing Emotions Accurately

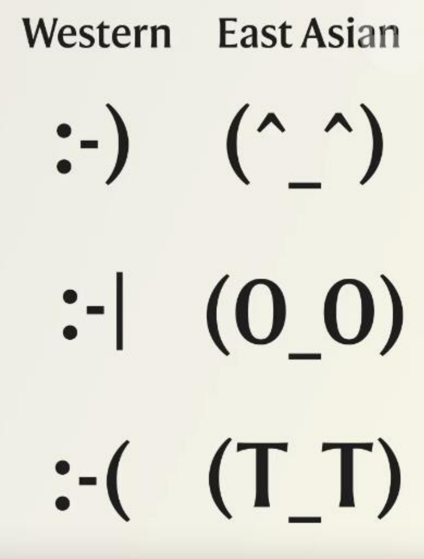
Since machine learning facial expression recognition (FER) systems have no innate understanding of human emotion, and since they rely on our ability to accurately label expressions, it is worth considering our own limitations in this respect, since these will only be magnified in a machine learning system built around them.

According[42] to one professor of psychology in the U.S., our own ability to interpret facial expressions amounts to much more than examining the topography of an isolated face (which constitutes the kind of data typically fed into FER systems).

> *"What we're dealing with is highly variable, high context-sensitive temporally changing patterns. While people are speaking, there are acoustical changes in the vocalizations, there are body postures. The person carries around with them a whole internal context of their body.*
>
> *"Then there's the external context: who else is present? What kind of situation are they in? What are they doing? ..so whereas scientists used to focus primarily on the face, now they're focusing more on faces in context."*

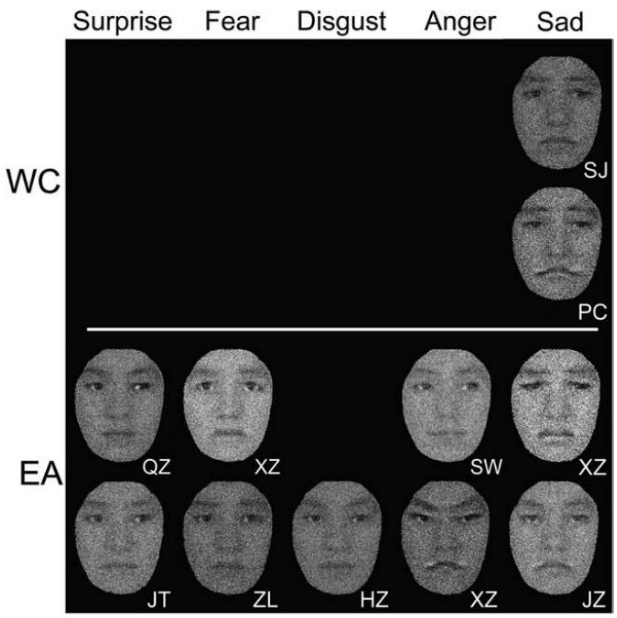
The Eyes Don't Always Have It

Research out of the University of Glasgow has observed that essential differences between the way that people from around the world manifest emotion is even reflected in the different characteristics of popular emoticons that one nation or region might prefer over another:



Emoticons developed in the West and in East Asia differ in the parts of the face that they emphasize. Western conventions assign emotional weight to the shape of the mouth, whereas East Asian equivalents emphasize the disposition of the eyes, and largely discount the mouth. Since these memes issue from equivalent behavior in the originating regions, machine learning FER systems will need racial context at the very least in order to identify emotions across a range of races, as well as across gender and age. Source: https://www.youtube.com/watch?v=sqOOezr32QM

A further study[43] on this topic pitted 15 white westerners against 15 East Asian subjects, with the aim of determining whether facial cues were equivalent across such distant nations, and finding that idiosyncrasies of expression do not cross borders well:



In East Asian culture, changing the direction of gaze is a central component of facial expression. In the row above, which represents the attempts of a White Caucasian WC group to correctly evaluate this trait in pictures of East Asian EA faces, we see that the WC group has only managed to identify 'sadness' out of the five gaze-averse emotions offered to them. Source: https://www.apa.org/pubs/journals/releases/xge-141-1-19.pdf

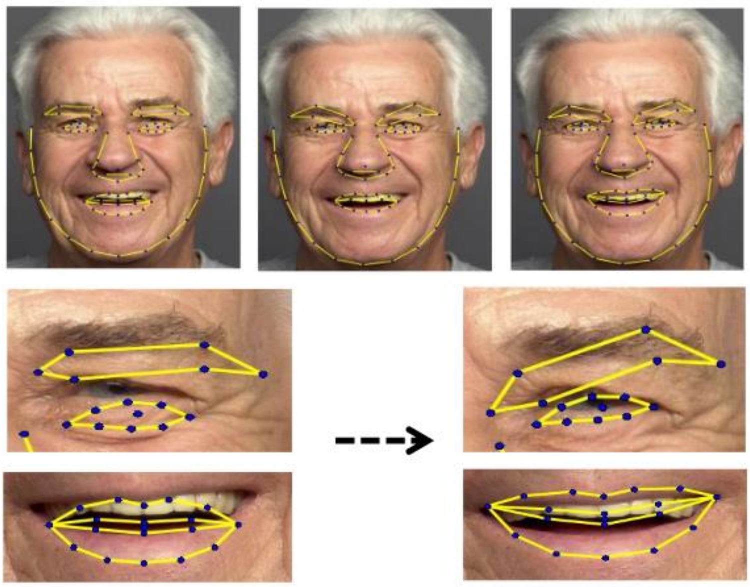
Taking Facial Expression Recognition Theory Beyond Ekman

A great deal of current thought and convention around FER has its roots in the 'facial action coding system'[44] developed by psychologist Paul Ekman in the 1970s. Ekman's theory assigned weight to 'microfacial' expressions deemed to be indicative of hidden emotions.

Though Ekman's methodology was widely adopted (and even became the direct inspiration[45] for the investigative drama *Lie To Me*), it has been the subject of growing scientific criticism[46] in recent decades, and its use as a terrorist-detection tool gauged at the same level as 'flipping a coin'[47].

In spite of this, Chinese authorities have deployed emotion recognition systems based on the same principles in the sensitive Xinjiang region in western China[48].

Even evaluating some of the supposedly less ambiguous expressions becomes problematic once we leave our own locality: a smile in a country beset by corruption can be interpreted as disingenuous, or even a sign of low intelligence[49]; facial expressions representing pain or sexual pleasure are quite different among diverse cultures[50]; the effects of ageing on the human face make expression recognition more difficult for us, and the machine systems we are informing[51]; and the labeling of expression data depends more on the interpretation of scientists than the direct feedback of participants[52] (who are the only ones who knew what they were feeling when the image was taken).

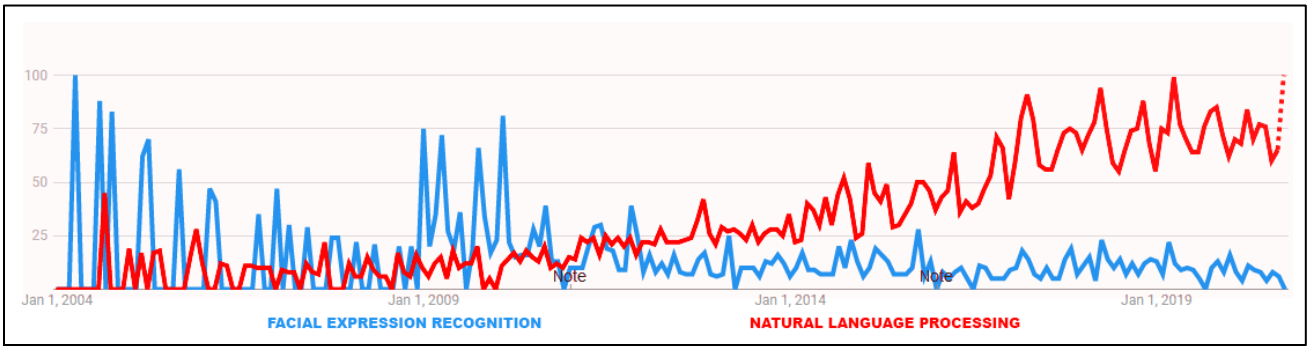


Research out of Iraq concludes that age identification is crucial to machine learning expression recognition systems. Here we see a Support Vector Machine (SVM) mistaking wrinkles for core facial features. Source: http://staff.cs.manchester.ac.uk/~tmorris/pubs/ICSIE_39_Nora_algaraawi.pdf

In addition to evidence that women in any one cultural group seem to use facial expressions differently to men[53], the well-supported[54] supposition that women are better able to correctly interpret facial expressions has potential implications for the data-gathering and pre-processing stage of any FER machine learning project.

Conclusion

Given the growing skepticism around Ekman-derived methodologies (and inter-related concerns regarding surveillance, governance and privacy), the sector's credibility problems seem to be defined more by doubt around its core assumptions than its machine learning-based implementations.



Google Trends between 2004-2020 for face-based and for linguistic emotion recognition search terms. Source: https://trends.google.com/

A historical glance at headlines around automated FER over the last decade or so reveals a number of bold announcements by variously-sized tech companies for FER products which either fail to materialize or else disappear, or are later downplayed by the originating company.

It would seem that this is the moment for proponents of facial expression recognition technologies to aim for performant products within a greater level of constraint and expectation, rather than continue to hope, against the growing evidence, that facial expressions represent a complete and discrete index of emotional response; and that social scientists and machine learning practitioners can entirely decode them, given enough time and resources.

It may be most beneficial to consider FER as a tool rather than an integral solution; to adopt multi-factor indicators of mood for more effective machine learning FER systems, adding diverse information streams around pose estimation, vocal tonality, and use of written language (where text-based user input is a factor).

If privacy and consent hurdles could be overcome and the use-case was compelling, the inclusion of health metrics such as estimated body temperature and blood pressure (via passive sensors or body-worn IoT devices) would also be valuable adjuncts in developing a more complete picture of emotional response.

---

[1] http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC1839771&blobtype=pdf

[2] https://jov.arvojournals.org/article.aspx?articleid=2138723

[3] https://www.sciencedaily.com/releases/2017/03/170313105901.htm

---

[4] https://www.salon.com/2019/06/03/cognitive-scientists-say-mona-lisas-smile-was-not-genuine/
[5] https://www.marketsandmarkets.com/Market-Reports/emotion-detection-recognition-market-23376176.html
[6] https://www.theguardian.com/technology/ng-interactive/2019/jun/22/the-rise-of-the-deepfake-and-the-threat-to-democracy
[7] https://www.ft.com/content/4bf4277c-f527-11e9-a79c-bc9acae3b654
[8] https://www.bbc.com/news/business-51204954
[9] https://openaccess.thecvf.com/content_CVPRW_2019/papers/Media%20Forensics/Agarwal_Protecting_World_Leaders_Against_Deep_Fakes_CVPRW_2019_paper.pdf
[10] http://www.ijitee.org/wp-content/uploads/papers/v9i6/E2779039520.pdf
[11] https://molecularautism.biomedcentral.com/track/pdf/10.1186/s13229-018-0187-7
[12] http://www.ijitee.org/wp-content/uploads/papers/v9i7/F3772049620.pdf
[13] http://autismglass.stanford.edu/
[14] https://pratt.duke.edu/about/news/amazon-autism-app-video
[15] https://digitalagenda.io/insight/artificial-intelligence-and-autism/
[16] https://www.ttnews.com/articles/some-fleets-reluctant-add-driver-facing-cameras
[17] https://gizmodo.com/your-next-car-will-be-watching-you-more-than-its-watchi-1840055386
[18] http://go.affectiva.com/in-cabin-sensing
[19] https://www.theguardian.com/media-network/2015/jul/27/artificial-intelligence-future-advertising-saatchi-clearchannel
[20] https://catalogofbias.org/biases/hawthorne-effect/
[21] https://www.telegraph.co.uk/news/2019/09/27/ai-facial-recognition-used-first-time-job-interviews-uk-find/
[22] https://www.inc.com/minda-zetlin/ai-is-now-analyzing-candidates-facial-expressions-during-video-job-interviews.html
[23] https://poseidon01.ssrn.com/delivery.php?ID=525111031127088064083094111096094124056042028006019024112093081117003120024092028070103101010014014024108094011025020099011091007025008015092027119066003080004110011019026047069095000027121073069073100079065079095010065123097107126016013112124008077000&EXT=pdf
[24] https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00208/full
[25] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3969247/
[26] https://www.technologyreview.com/2019/11/07/75194/hirevue-ai-automated-hiring-discrimination-ftc-epic-bias/
[27] https://www.washingtonpost.com/technology/2019/11/06/prominent-rights-group-files-federal-complaint-against-ai-hiring-firm-hirevue-citing-unfair-deceptive-practices/
[28] https://tipsmake.com/leading-companies-in-korea-use-ai-to-recruit-employees-and-netizens-to-find-ways-to-cope
[29] https://www.scmp.com/news/asia/east-asia/article/3045795/south-korean-job-applicants-are-learning-trick-ai-hiring-bots
[30] https://www.theverge.com/2014/10/10/6954135/barcelona-comedy-club-is-charging-customers-by-the-laugh
[31] https://skybiometry.com/documentation/
[32] https://github.com/microsoft/Cognitive-Emotion-DotNetCore
[33] https://cloud.google.com/vision/docs/features-list
[34] https://docs.aws.amazon.com/rekognition/latest/dg/API_Emotion.html
[35] https://cloudacademy.com/blog/google-vision-vs-amazon-rekognition-a-vendor-neutral-comparison/
[36] https://patentimages.storage.googleapis.com/fe/f5/d5/76c91c50749b0c/US8209182.pdf
[37] https://www.pocket-lint.com/apps/news/apple/142230-what-are-animoji-how-to-use-apple-s-animated-emoji
[38] https://9to5mac.com/2016/01/07/apple-emotient/
[39] http://appft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahtml%2FPTO%2Fsearch-adv.html&r=16&p=1&f=G&l=50&d=PG01&S1=(apple.AANM.+AND+20191114.PD.)&OS=aanm/apple+and+pd/11/14/2019&RS=(AANM/apple+AND+PD/20191114)
[40] https://appleinsider.com/articles/19/11/14/future-versions-of-apples-siri-may-read-interpret-your-facial-expressions
[41] https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0231968&type=printable
[42] https://www.youtube.com/watch?v=sqOOezr32QM

[43] https://www.apa.org/pubs/journals/releases/xge-141-1-19.pdf
[44] https://www.paulekman.com/facial-action-coding-system/
[45] https://www.ekmaninternational.com/about/lie-to-me/
[46] https://www.bostonmagazine.com/news/2013/06/25/emotions-facial-expressions-not-related/
[47] https://www.nature.com/news/2010/100526/full/465412a.html
[48] https://www.ft.com/content/68155560-fbd1-11e9-a354-36acbbb0d9b6
[49] https://www.theatlantic.com/science/archive/2016/05/culture-and-smiling/483827/
[50] https://www.pnas.org/content/115/43/E10013
[51] http://staff.cs.manchester.ac.uk/~tmorris/pubs/ICSIE_39_Nora_algaraawi.pdf
[52] https://www.technologyreview.com/2019/07/26/238782/emotion-recognition-technology-artifical-intelligence-inaccurate-psychology/
[53] https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0173942&type=printable
[54] https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0190634&type=printable