

The Future of Emotion Recognition in Machine Learning

Originally published at <https://www.iflexion.com/blog/emotion-recognition-software>

Published: October 21, 2020 | By Martin Anderson



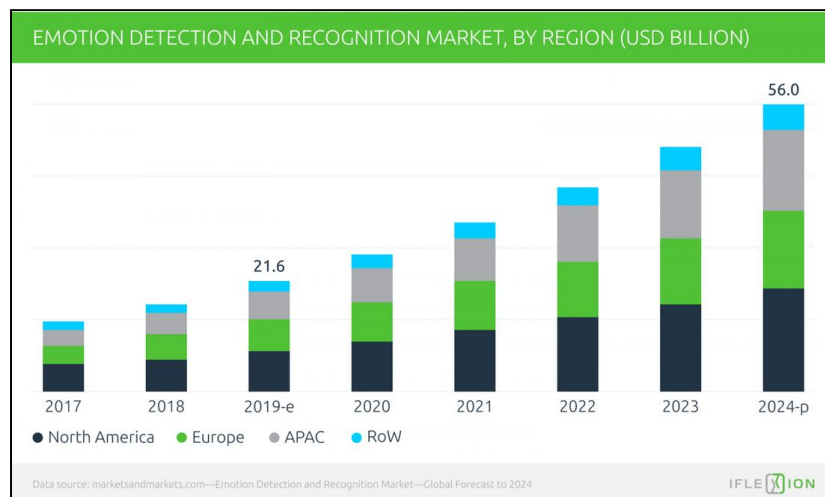
Understanding the meaning of facial expressions is an essential human survival tool, but one that generates scientific controversy. Some contend that we recognize anger most easily¹, since it may represent an existential threat; others that we recognize happiness most easily and anger most slowly², since 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 remain split even on whether the Mona Lisa's enigmatic smile is sincere³ or forced⁴ (apparently a matter of context).

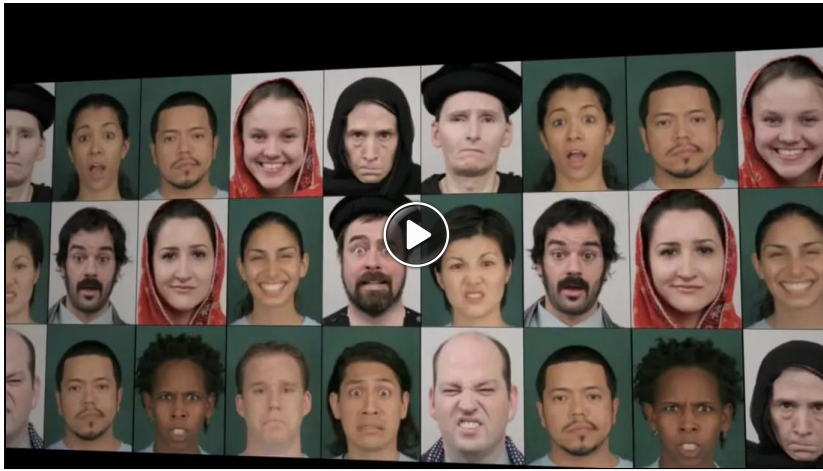
The ability to decipher the true intent and emotional response of a person from their facial expressions, notwithstanding their attempts to mask or deceive what they feel, is an evolutionary advantage of great interest to a range of sectors, from physicians through to marketers and political analysts.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

Market Growth of Emotion Recognition Software

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 2024⁵.



The advent of industrialized machine learning techniques has combined with the 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 the 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](#)?

Is facial expression recognition a solution in search of a problem, really?

Uses for Expression Recognition

Deepfake Detection

Academic research into identifying deepfake video manipulations ramped up in the 12 months prior to the 2020 US 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 subjects⁹ — the only one, among several novel approaches¹⁰, that relies specifically on facial expressions.

DEEPPFAKE DETECTION

From left to right: the target subject, impersonator and 'deepfaked' impersonator.



Data source: openaccess.thecvf.com—Protecting World Leaders Against Deep Fakes, 2019

IFLE 

Because it is more difficult to obtain a high volume of images of 'rarer' expressions for a machine learning dataset, 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 emotions¹¹.

One 2018 study¹² 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:

HAAR CASCADE OBJECT DETECTION IN ASD RESEARCH AND THERAPY



Data source: [jitsee.org](https://github.com/jitsee.org)—Facial Emotion Recognition for Autism Children, 2020

IFLE 

Stanford University's Autism Glass Project¹³ uses Google's face-worn computing system to aid autism-affected subjects in understanding appropriate social cues by appending emoticons to those expressions the system is able to recognize.

THE AUTISM GLASS PROJECT: LABELING FACIAL EXPRESSIONS WITH EMOTICONS



Data source: [youtube.com](https://www.youtube.com/watch?v=...)—Google Glass helps kids with autism read facial expressions

IFLE 

Another project¹⁴ 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](https://pytorch.org/), and TensorFlow.

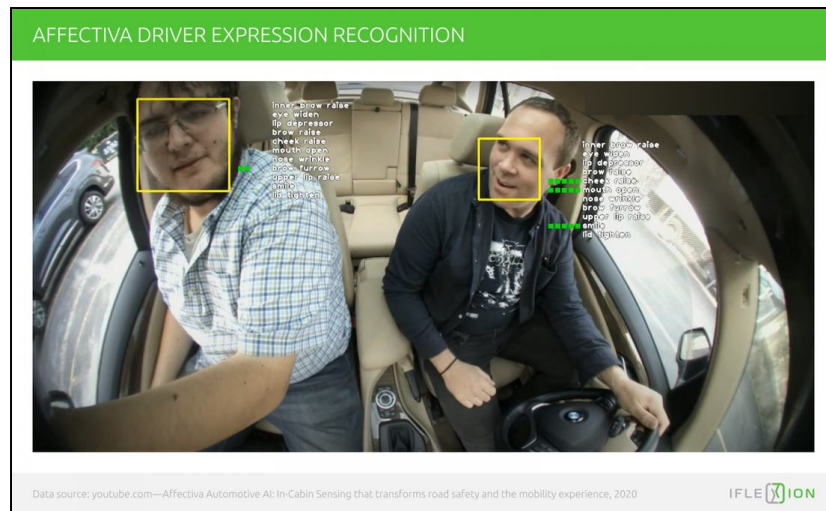


Since autism is such an apt candidate for expression recognition technologies, there are more applications of machine learning for this cause¹⁵ than we are able to list here.

Automotive Safety and Research Systems

Due to increased controversy over data mining and privacy concerns¹⁶ 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 anticipation¹⁷ 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 includes an expression-recognition component that's capable of determining if the driver is falling asleep.

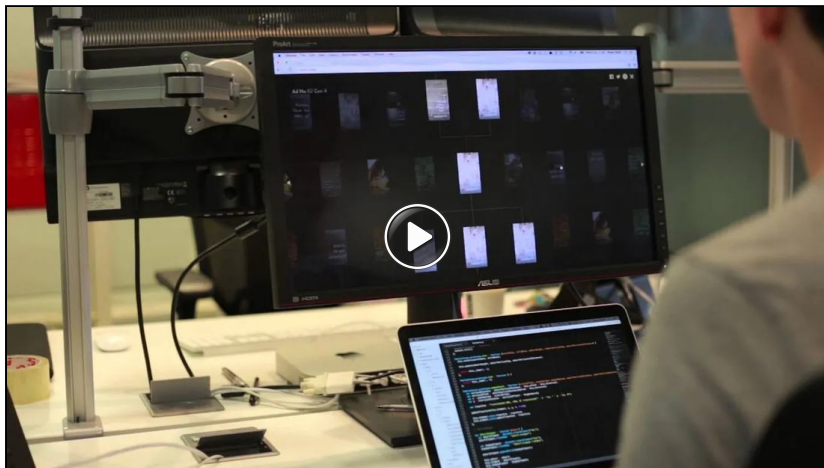


This test for 'liveness' or alertness can be configured to trigger a number of warnings or other safety-related actions, which is particularly relevant as a driver control feature in [commercial fleet management software](#).

Market Research

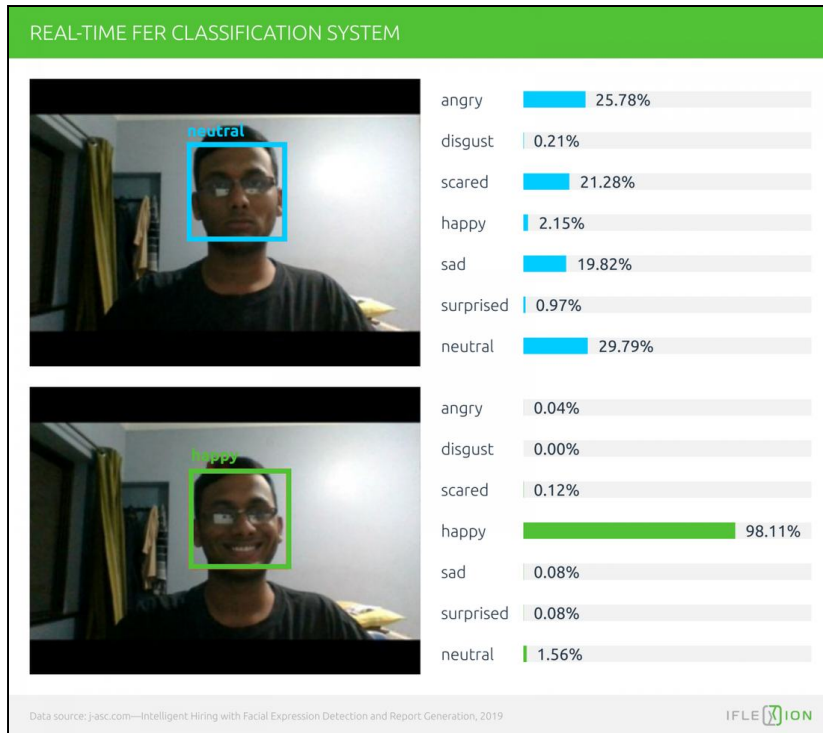
Since monitoring user data by stealth has become such a political football in recent years, facial expression recognition technologies are on their strongest ground where the subjects are willing accomplices, such as in focus groups and other forms of beta-testing for product marketing.

Outside of such controlled environments, there are possibilities for novel use of FER in a marketing context, such as M&C Saatchi's 2015 experiment¹⁸ with an AI-powered poster that changed the advertising content depending on the expressions that it could read from passers-by.



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 effects¹⁹ 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 and thus hindering [intelligent video analysis](#).



Though FER is being increasingly implemented in a recruitment context^{20,21}, it comes with a number of caveats:

- Depending on the system, it can introduce racial bias²² that can expose employers to legal repercussion (a sub-set of wider problems around ethnicity in facial recognition²³).
- It will probably work best without user consent²⁴ — usually illegal in the EU and in many other jurisdictions.
- It seems set for further regulation^{25,26}, 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 recruiting](#)²⁷, candidates are now schooling themselves²⁸ 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 shows²⁹, installing emotion recognition cameras in front of each seat, and fixing a ceiling of 24 euros per customer for the more successful events.



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.

From deepfake detection to autism therapy research, there are multiple emerging use cases for facial expression and emotion recognition right now.

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 API³⁰, which provides a range of facial detection and analysis features, can also individuate anger, disgust, neutral mood, fear, happiness, surprise and sadness. Microsoft Azure's Emotion API³¹ 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 Vision³² and Amazon Rekognition³³, both of which provide facial sentiment detection facilities as a component of their more broadly successful facial recognition APIs.

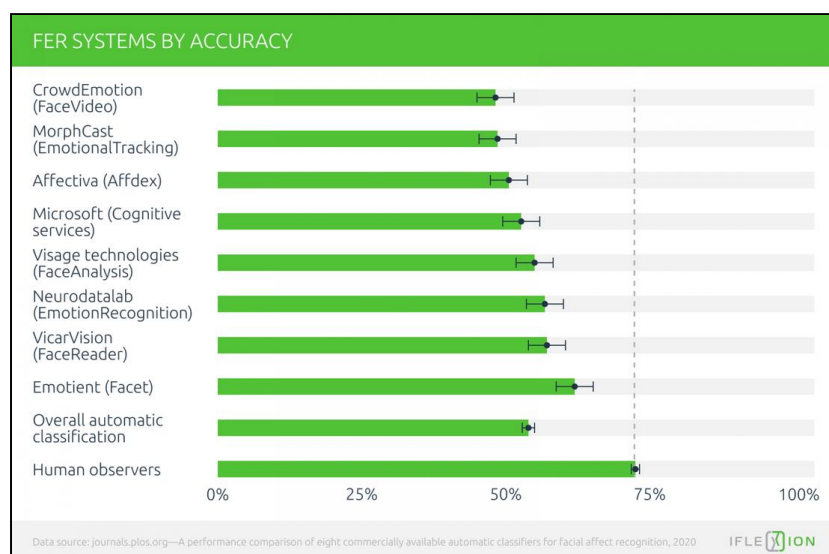
Though their comparative offerings are not identical in terms of properties or methods of quantifying success, one 2019 comparison³⁴ of the services found that Google was less likely to commit to identifying an emotion at all, whereas Rekognition is willing to commit to an emotion at levels as low as 5% in order to return a result of some kind.

Emotient

Apple was patenting emotion recognition mechanisms as far back as 2012³⁵. However, besides an expression-based emoji system³⁶, the tech giant has delivered little working FER functionality in the wake of its acquisition of artificial intelligence startup Emotient in 2016³⁷.

However, a more recent 2019 patent³⁸ has led to a speculation³⁹ 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 study⁴⁰ compared eight commercially available FER systems against human responses to facial expressions, with Apple's Emotient system scoring highest.



The Challenge of Recognizing Emotions Accurately

Since machine learning 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 to one professor of psychology in the US⁴¹, 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.

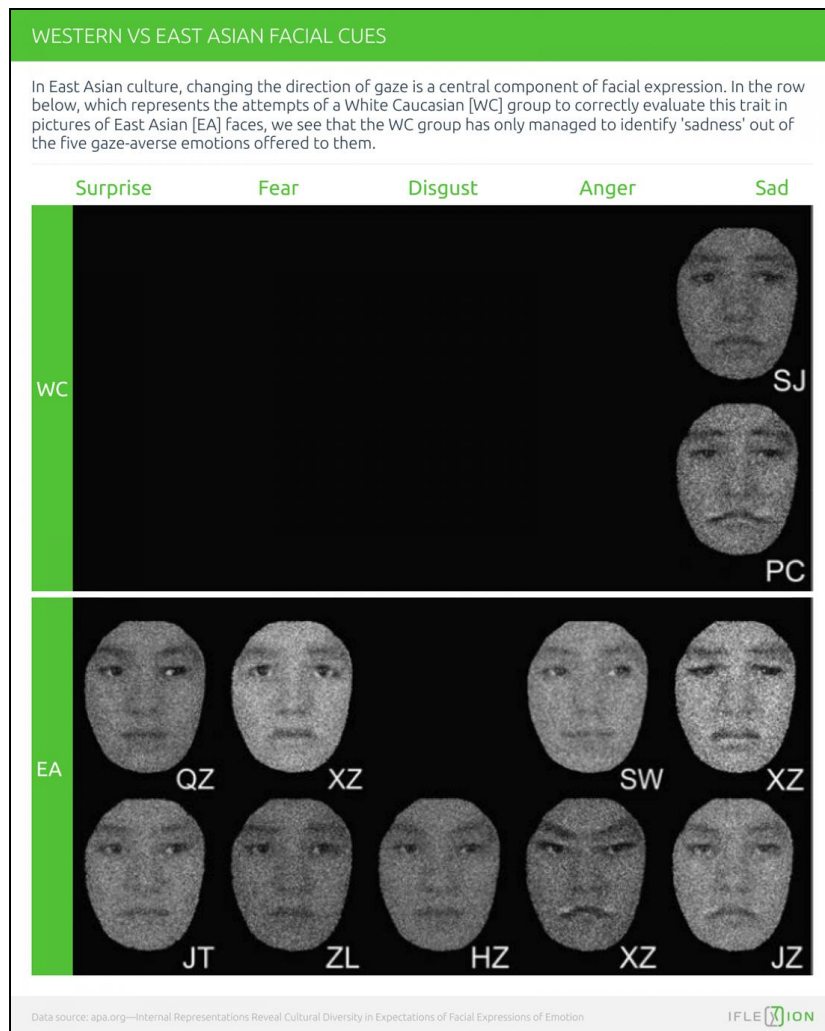
Lisa Feldman Barret, Professor of Psychology, Northeastern University

”

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:

A further study⁴² 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:



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'⁴³ developed by psychologist Paul Ekman in the 1970s. Ekman's theory assigned weight to 'microfacial' expressions deemed to be indicative of hidden emotions.

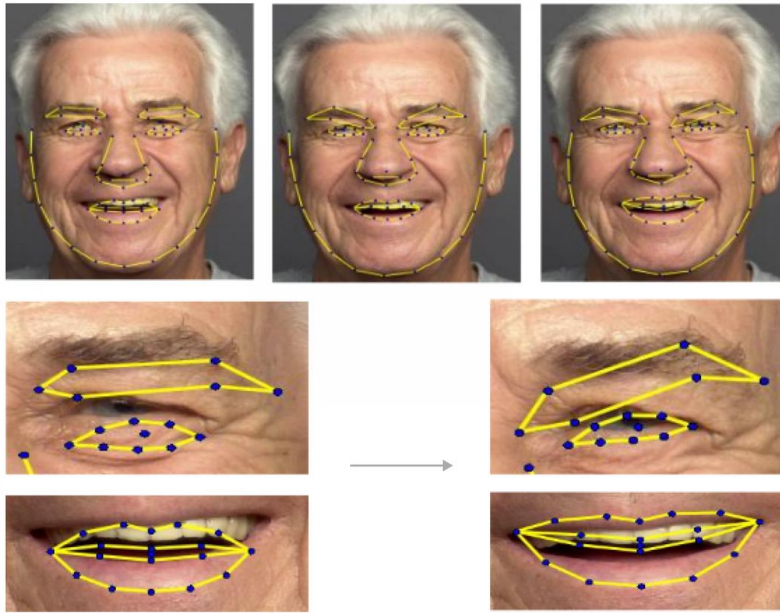
Though Ekman's methodology was widely adopted (and even became the direct inspiration⁴⁴ for the investigative drama *Lie To Me*), it has been the subject of growing scientific criticism⁴⁵ in recent decades, and its use as a terrorist-detection tool gauged at the same level as 'flipping a coin'⁴⁶.

In spite of this, Chinese authorities have deployed emotion recognition systems based on the same principles in the Xinjiang region in western China⁴⁷.

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 intelligence⁴⁸; facial expressions representing pain or pleasure are quite different among diverse cultures⁴⁹; the effects of ageing on the human face make expression recognition more difficult for us and the machine systems we are informing⁵⁰; and the labeling of expression data depends more on the interpretation of scientists than the direct feedback of participants⁵¹ (who are the only ones who knew what they were feeling when the image was taken).

AGING FACTOR IN EXPRESSION RECOGNITION

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.



Data source: staff.cs.manchester.ac.uk—Study on Aging Effect on Facial Expression Recognition

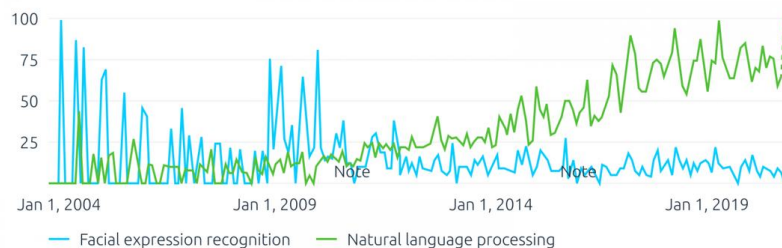
IFLE ION

In addition to evidence that women in any one cultural group seem to use facial expressions differently to men⁵², the well-supported supposition that women are better able to correctly interpret facial expressions⁵³ has potential implications for the data-gathering and pre-processing stages of any FER machine learning project.

Conclusion

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

INTEREST IN SENTIMENT ANALYSIS AND FACIAL EXPRESSION RECOGNITION, 2004—2020



Data source: Google Trends

IFLE ION

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, and 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.

This is the moment for proponents of facial expression recognition technologies to aim for performant products rather than continue to hope that facial expressions represent a complete and discrete index of emotional response.