

## AI Vision

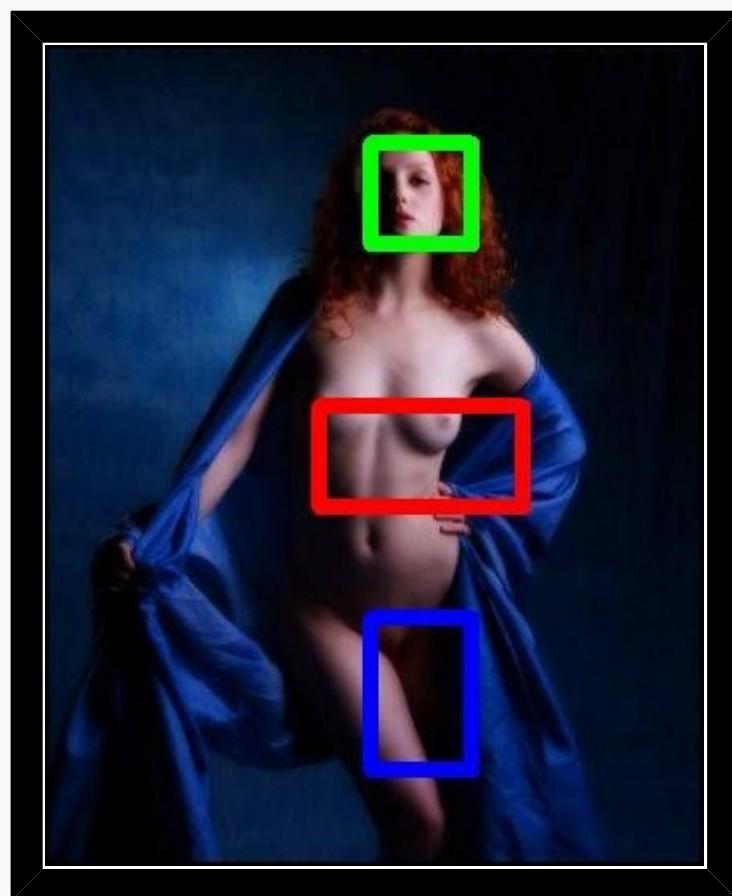
# Can off the shelf AI Vision systems detect and censor art nude photographs? A comparison test of Google Vision and Microsoft Vision

### Summary:

Question: can AI vision systems from Microsoft and Google, which are available for free to anybody, identify NSFW (not safe for work, nudity) images? Can this identification be used to automatically censor images by blacking out or blurring NSFW areas of the image?

Method: I spent a few hours creating in some rough code in Microsoft office to find files on my computer and send them to Google Vision and Microsoft Vision so they could be analysed.

Conclusion: Yes. I spent a few hours over the weekend just knocking some very rough code. Yes, they did reasonably well at (a) identifying images that could need censoring and (b) identifying where on the image things should be blocked out.



(<http://www.filmphotoacademy.com/blog/X4.jpg>)

Automatically redacted zones

Follow on question: Why aren't sites like Facebook and Instagram automatically deploying this technology to identify images and allowing users to choose whether they wish to see such images?

How do we know how much of our internet is already censored invisibly?

**Links to downloads of the image results and the code I used can be found at the end of the article**

## The problem:

As an art nude and fashion photographer one of my persistent problems is determining what can or cannot be posted to various social media sites. On the whole if it is “non FaceBook friendly” as we describe it then I just don't post it. However, if you shoot nudes and fashion it is certain that there are pictures you want to show off – or encourage viewers to view in full on your website. Also, many, many of us photographers have fallen foul of the FaceBook police even when we think we are being careful.

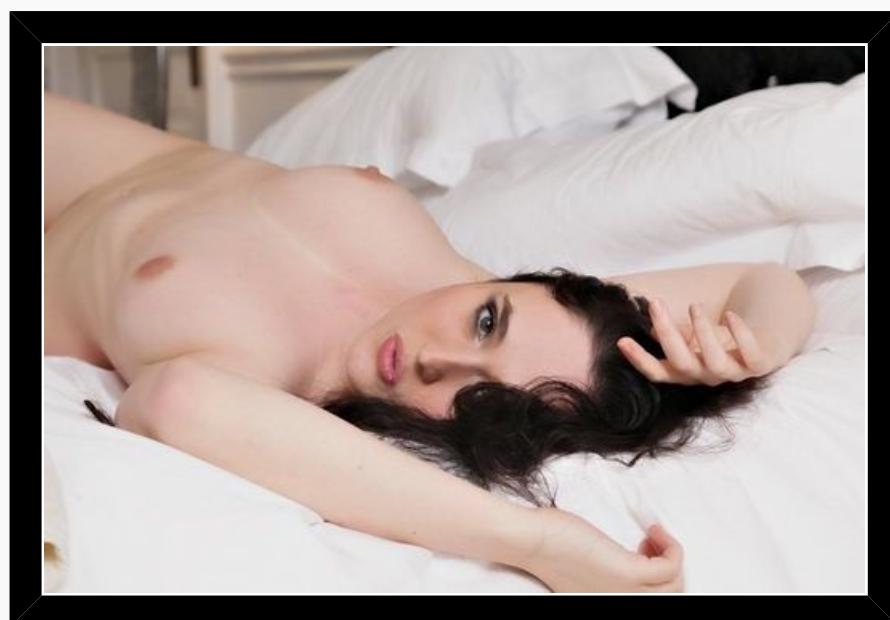
I have tried hand editing NSFW images for posting to FaceBook and Instagram, but it is very tedious and time consuming. So I wanted to know if I could automate the process. I was also curious about how effective the key AI vision systems were.

And if you are running a photography site that allows posting of images, then you have a real nightmare on your hands – you rely on posters tagging their images correctly, but what about the many who forget? These systems do provide a fairly simple system for automatically checking images for further review or other action.

## The method:

I processed 4,300 art nude, fashion, portrait and random images through the Google Vision and Microsoft Vision artificial intelligence systems to see how good they were. I put together about 200 lines of code in Microsoft Office to find files and make AI requests. This was all rough and ready, not meant to be a finished or polished system but to give me an idea of the capabilities.

Firstly, their actual identification for NSFW wasn't as good as I thought it ought to be – with a 10% failure rate. I fed in around 4,300 images. Between the two systems they rated 1882 as “safe”. Of these 167 had at least one clearly visible breast with nipples, and often a lot more nudity than that. So give or take 10% failed allowing clearly nude images into the “safe” zone.

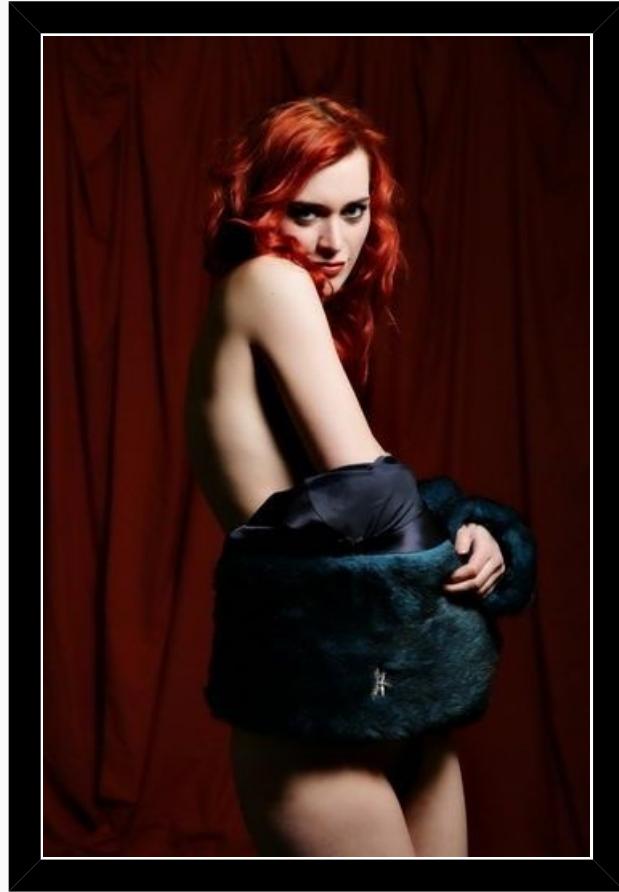


([http://www.filmphotoacademy.com/blog/X\\_FAIL4.jpg](http://www.filmphotoacademy.com/blog/X_FAIL4.jpg))

AI failed to recognise as NSFW!

I'm not fazed or upset by scantily clad women or sexualised situations. But these systems weren't great at recognising the rude amongst the safe.

I would say easily 30% of the "safe" images I would class as "racy" – and this is my genre



([http://www.filmphotoacademy.com/blog/X\\_FAIL6.jpg](http://www.filmphotoacademy.com/blog/X_FAIL6.jpg))

AI rates as "safe" when definitely "racy"

OK, so how did it do on the "racy" side. Pretty well to be fair. I would probably rate 20 images as actually being safe, but would understand why they were selected as racy. All of these were images where cleavage was a feature but no bust showing. So few false positives.



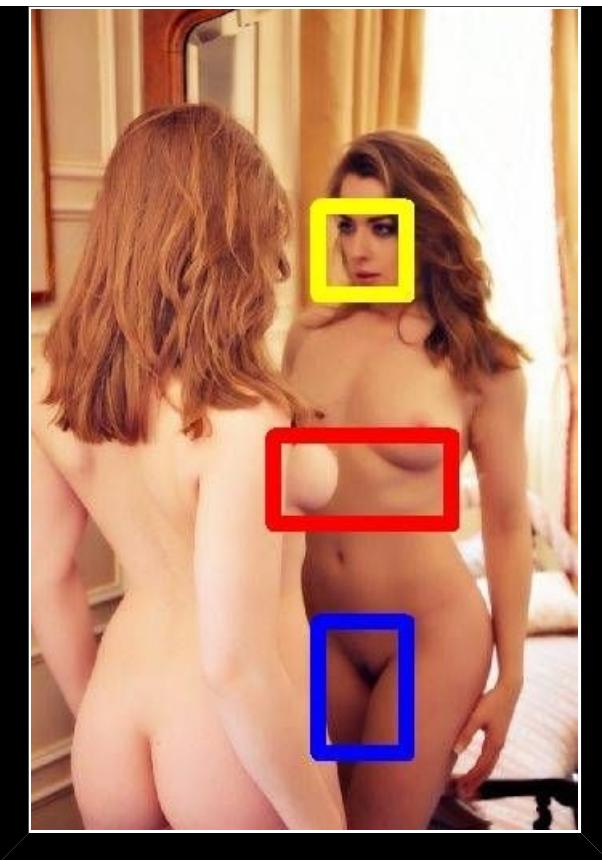
([http://www.filmpphotoacademy.com/blog/X\\_FAIL10.jpg](http://www.filmpphotoacademy.com/blog/X_FAIL10.jpg))

AI systems classes as “racy”

I should add, that these systems return a confidence rating along with a Yes/No response. You could choose to play it safer. In turn though that would increase the number of false positives as well.

## From AI to my redaction attempts

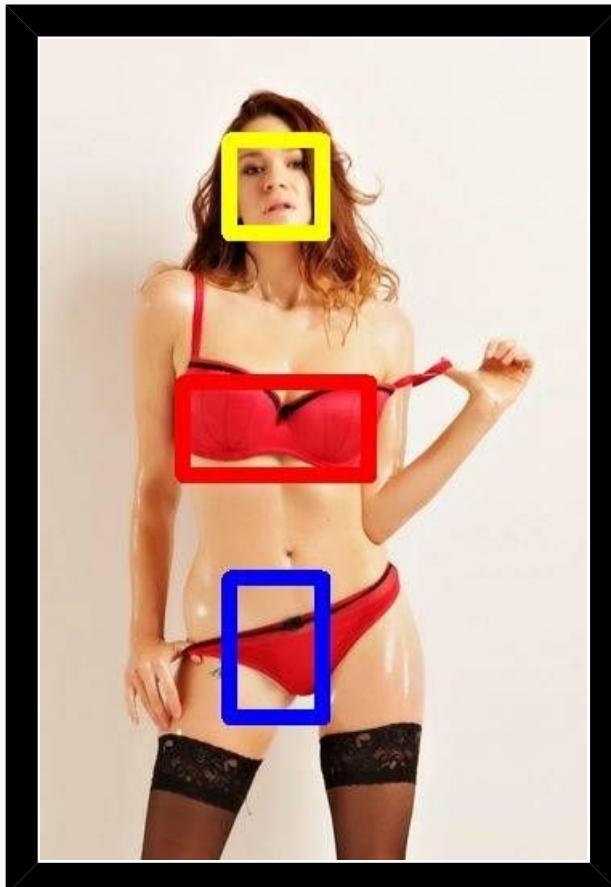
Back to my censoring attempts, I took a very simple approach, just assume that boobs and pubes are directly below the face. This is where you would usually expect to find them! My effective rate wasn't bad.



(<http://www.filmphotoacademy.com/blog/X1.jpg>)

Redacted Zones pretty close

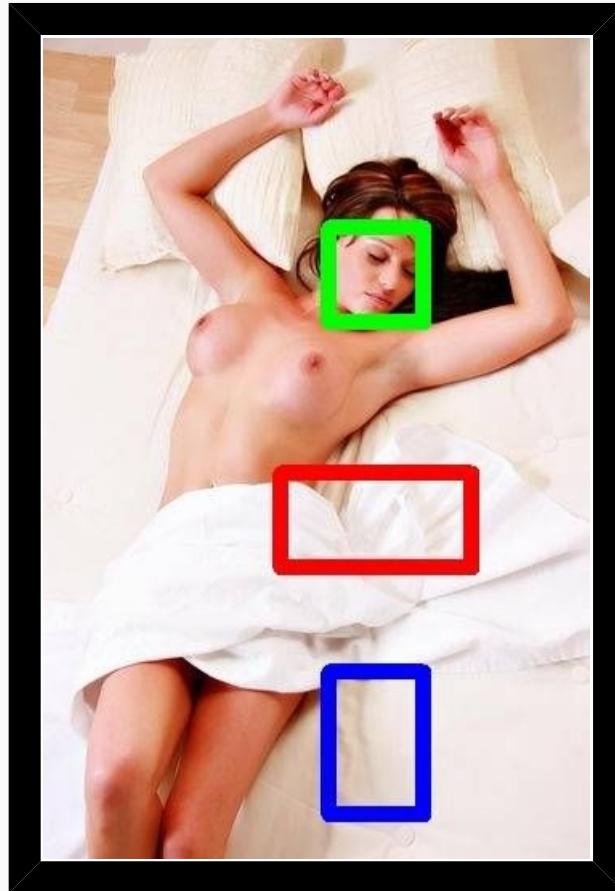
However, I was unable to distinguish between nudes and underwear. Hence masking bikini tops and bottoms.



(<http://www.filmphotoacademy.com/blog/X2.jpg>)

## Redacts clothed figures

My test system could not cope with the body being posed at an angle. Both systems could also be thrown off with faces at an angle, they would give quite different face size estimates. This in turn throws my estimations off. Google does provide significant information about the tilt, roll and yaw of the face – so a little more effort could possible accommodate that.

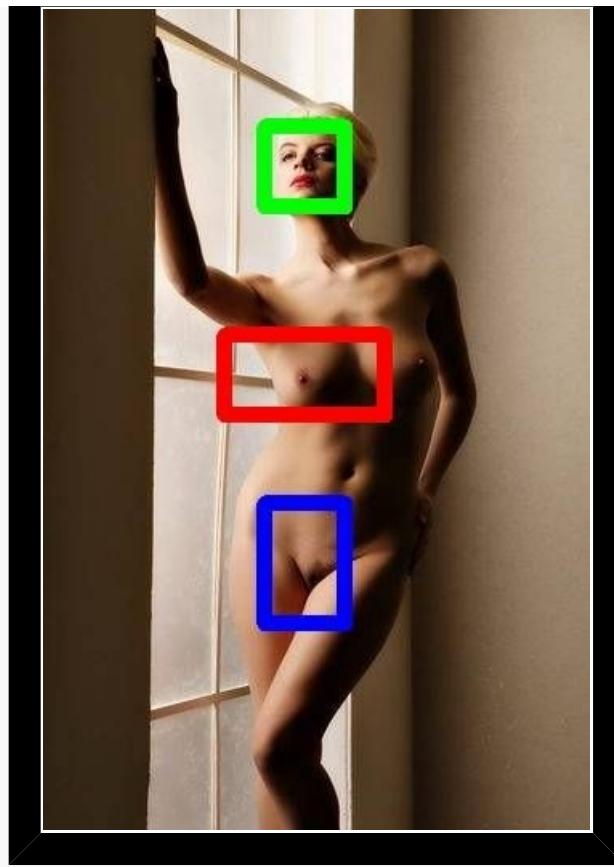


([http://www.filmphotoacademy.com/blog/X\\_FAIL3.jpg](http://www.filmphotoacademy.com/blog/X_FAIL3.jpg))

Cannot handle figures at an angle

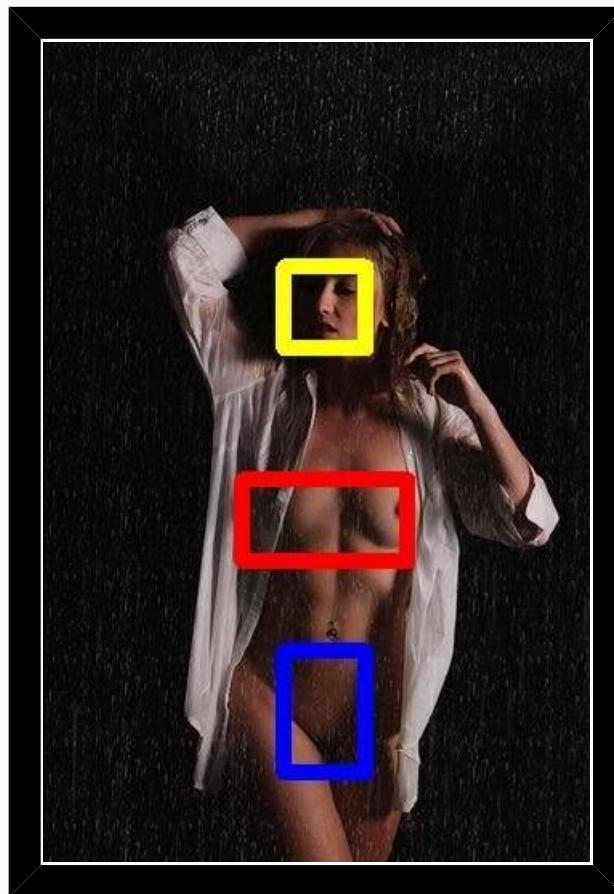
Finally, you are left with the question – how do you apply the censorship. Neither system would let me target nipples or genitals directly. So I was left with a bit of a crude rectangular box as the answer. I am sure – given how much detail about a face it is possible to get it would be trivial to train an AI to spot nipples and genitals. With such points you could then choose to blur very selective regions as pass the requirements for sites that filter “adult” images.

With a little more coding effort I suspect you could just search for the darkest areas within the centre of the bounding box and you would be right.



([http://www.filmpphotoacademy.com/blog/X\\_FAIL1.jpg](http://www.filmpphotoacademy.com/blog/X_FAIL1.jpg))

Redaction confused by standard feminine "S" pose



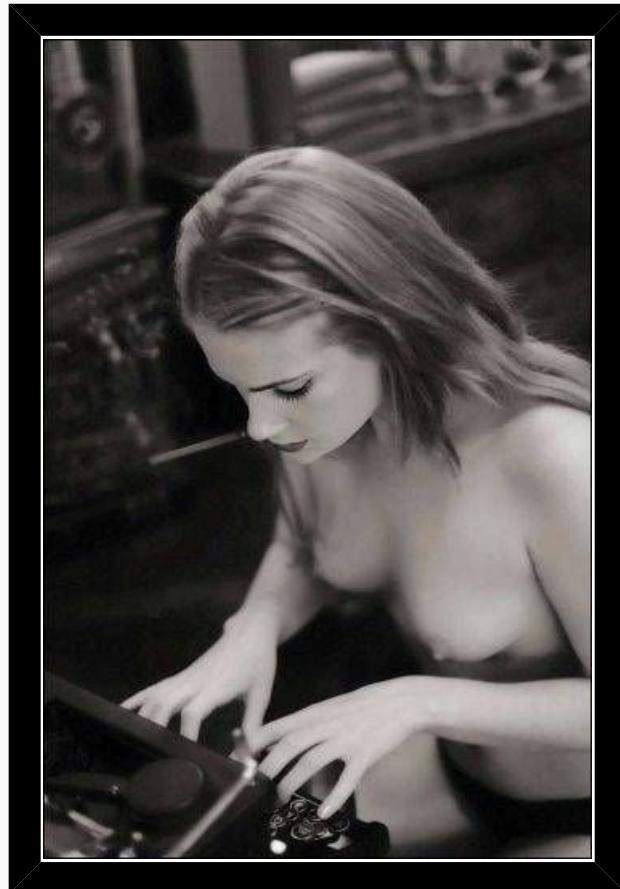
(<http://www.filmpphotoacademy.com/blog/X3.jpg>)

Redaction precise and fairly minimal.

## Other redaction problems?

It relies on finding a face. If there is no face in the picture then it can do nothing. So images where the head is cropped or images of the back of a person will get through every time.

But even with a face in view, if it is not recognised – which is surprisingly frequent – then no redaction can take place.



([http://www.filmphotoacademy.com/blog/X\\_FAIL9.jpg](http://www.filmphotoacademy.com/blog/X_FAIL9.jpg))

AI systems failed to find face



([http://www.filmpphotoacademy.com/blog/X\\_FAIL8.jpg](http://www.filmpphotoacademy.com/blog/X_FAIL8.jpg))

No face visible for the AI to recognise

## Artificial Censorship and You

So, now we know that computers can do the censoring, with a high degree of flexibility and targeting. In fact, they can do more than censor nudity. I discovered the Google Vision system for example can identify violent and medical images – it will even detect something called “spoof” images- which in these days of fake news is also interesting.

This gives us three immediate questions are (a) are they doing it now? (b) how would you know? and (c) should they?

### Are they doing it now?

Yes, an obvious example is Google's “safe search” on images.



([http://www.filmphotoacademy.com/blog/TITS\\_SAFE.JPG](http://www.filmphotoacademy.com/blog/TITS_SAFE.JPG))



([http://www.filmphotoacademy.com/blog/TITS\\_NOSAFE.JPG](http://www.filmphotoacademy.com/blog/TITS_NOSAFE.JPG))

Google Search for "tits", not filtered compared to filtered.

How about FaceBook. At the moment FaceBook says that they rely on user reports. If that is really true I am amazed! Given that my simple tests show they could automatically reduce NSFW content by 90%. Which is also odd, since FaceBook very strongly manages what you see in your newsfeed. If you didn't like art nudes you would think they wouldn't appear?

## How would you know?

This is much trickier. By definition, you don't know what you don't know. You may think you are seeing everything, but you cannot determine what might have been filtered before it gets to you.

We used to live in a world where newspapers and TV determined what we saw – and we knew deep down that they only showed us what we wanted to see. But with the internet we thought we had access to everything – our news and information were no longer biased and filtered. Well, this year we certainly found out that wasn't the case. Our information world is completely algorithmically generated. We only know we don't know if a friend tells us – and probably they don't know either!

## Should it be filtered at all?

I am really strongly against filtering at root source and very strongly in favour filtering at the consumer end. If I have children then of course I want some subjects filtered. If I am a grown responsible adult I can make my own mind up about what I want to see and read.

I want to protect children from violence, I don't think they need protecting from bare breasts. So my personal preferences may not be yours. If you were a vegan you might choose not to show animal abuse, or you might want to inspire you to take action.

So, yes I would like a world where I could say, "yep, tell me how Aleppo or Palestine are – but don't show me really violent pictures, on the other hand, I don't want to see porn but I would like to see art photography."

However, this may become a bit moot – especially in the UK where the government is pressing to block sites with "adult content", including female orgasms and menstruation. Of course that "adult content" is what you choose it to mean. I can think of a wide range of art nude photographers who could easily be classed as pornographic depending on the viewer.

## How does it all work?

Both Google and Microsoft Vision products have a machine learning system behind them which have been taught to identify many items within photographs. This could be anything, from landscapes, buildings, still life, interiors and people. Now obviously with more photos of people than anything else there is a concentration of effort here.

For smaller numbers of images these services are available off-the-shelf and free to use. This is the part I was interested in. Could a free service, completely untrained for the specifics of what I want be able to do the job?

Both systems will attempt to identify faces – which they do OK, I was expecting better. When a face is identified then you can even drill down to find the eyes, ears, nose, and mouth at great detail.

Both systems will attempt to identify the image as Adult for Google and Adult or Racy for Microsoft. I combined these and assumed if one of them triggered then it was adult.

My crude attempt uses both systems to find the face. Once the face is identified then I just mark out areas at specific distances down the body which could be bust or pubes. There's nothing clever here at all.

I have a strong suspicion that both of these systems are capable of doing much more, but they are not making that facility publicly available. Firstly, it would represent a huge market opportunity to other companies needed to provide more sophisticated filtering services. Secondly, every porn watcher who could hack some code could have their porn pre-filtered for "only the good stuff".

I should repeat – these are off-the-shelf. They are not trained specifically for nudity but as a general purpose tool.

I am an experienced coder, but not familiar with either of the two systems, but it took me about 4 hours of coding to get a system running that could feed any image and get the response back.

## So what else can Artificial Intelligence tell us?

A side interest I had to all this was whether I could get useful image description or tags for the images. I have been extremely lazy and none of my work is keyworded in anyway at all.

Google provides a list of keywords and some info about emotions possibly present in the image. Microsoft doesn't do the emotions but does try to create a descriptive caption for the image. Microsoft also specifically tries to guess gender and age of the person. Microsoft does have a more focussed emotion system, but as a different interface and was not tested.

## Does it work?

No. Not really and not well enough. It is a bit more like reading your star signs in the newspaper – make it vague enough and throw enough into the mix and some of it will be right.

Microsoft seems to think a lot of images include water, umbrellas, dogs and cellphones. For example, water appears in 25% of all tagged images – I have no idea why. Cell phones I sort of understand – it seems to be reading them from the typical model poses of hand to face.

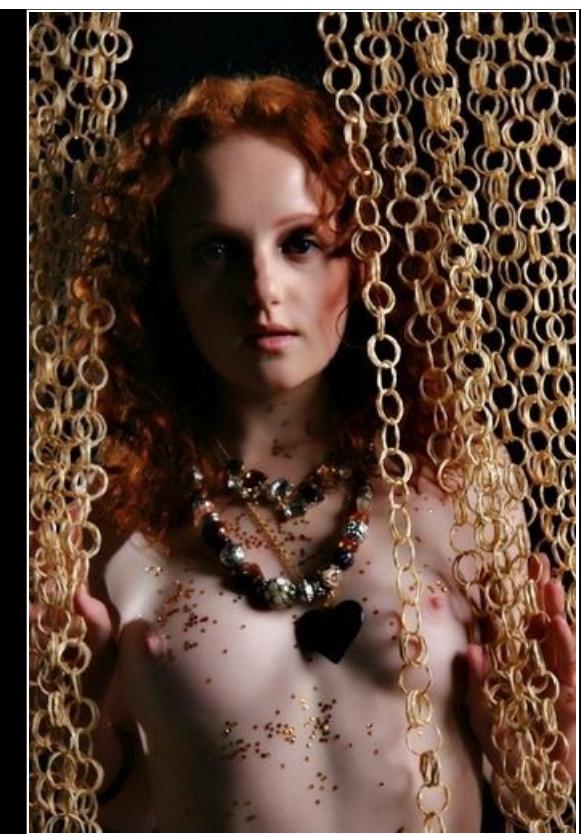
Microsoft also thinks a lot of my subjects are brushing their teeth or playing tennis. Again I think this is the AI reading the pose and trying to guess the activity.



([http://www.filmpphotoacademy.com/blog/Art\\_LastSupper-1024x424.jpg](http://www.filmpphotoacademy.com/blog/Art_LastSupper-1024x424.jpg))

Microsoft: a group of people are brushing their teeth

And my personal favourite



([http://www.filmphotoacademy.com/blog/~03-SQW-OLD~2007-WORK~Holly-2007-06-06\\_HR\\_2180.jpg](http://www.filmphotoacademy.com/blog/~03-SQW-OLD~2007-WORK~Holly-2007-06-06_HR_2180.jpg))

Microsoft: A woman looking at a giraffe

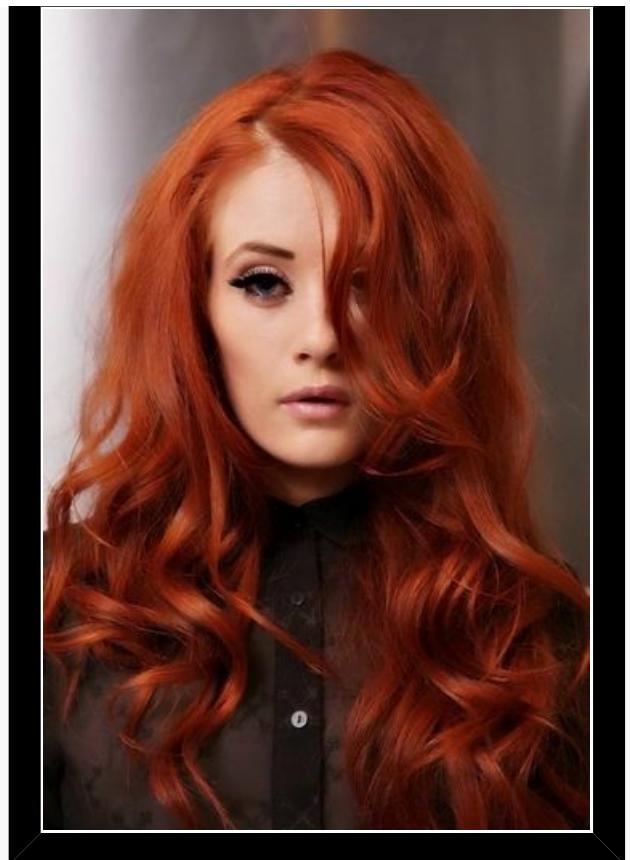
Not surprisingly, and on the whole quite accurately many of my images were correctly captioned as some variation of “a woman posing for a picture” even with non-standard image compositions, for example this one of a model lying down.



(<http://www.filmphotoacademy.com/blog/Flood-06C.jpg>)

Microsoft: A woman posing for a picture

Google does not provide captions, but its keywords seem much more spot on. Google returns fewer keywords (this can be controlled and I asked for 10 results per image, the default seems to be 3 to 5). So for this image fewer keywords but very specific to the image



([http://www.filmpphotoacademy.com/blog/XP\\_Hair2.jpg](http://www.filmpphotoacademy.com/blog/XP_Hair2.jpg))

Google tags: Hair, human, hair color, color, red, face, blond, hair-coloring, hairstyle, long-hair, brown-hair



([http://www.filmpphotoacademy.com/blog/XP\\_Adventure.jpg](http://www.filmpphotoacademy.com/blog/XP_Adventure.jpg))

Google tags: Clothing, woman warrior, costume,

Interesting Microsoft identified this a “woman wearing a hat” but returned this rather curious mix of tags “wearing,dressed,dress,standing,girl,brown,red,hat,umbrella,bear,room” – where's the bear!!

If we look at a couple of specific examples:



([http://www.filmphotoacademy.com/blog/jm\\_1780.jpg](http://www.filmphotoacademy.com/blog/jm_1780.jpg))

Microsoft: a beautiful young woman holding a frisbee

Microsoft tags: female, girl, beautiful, dress, beach, standing, playing, tennis, wearing, court, player, top, frisbee, water, air, mouth, board, ball, catch, game

Google tags: dance, performing arts, sports, modern dance, concert dance, muscle, leg, ballet, physical fitness,



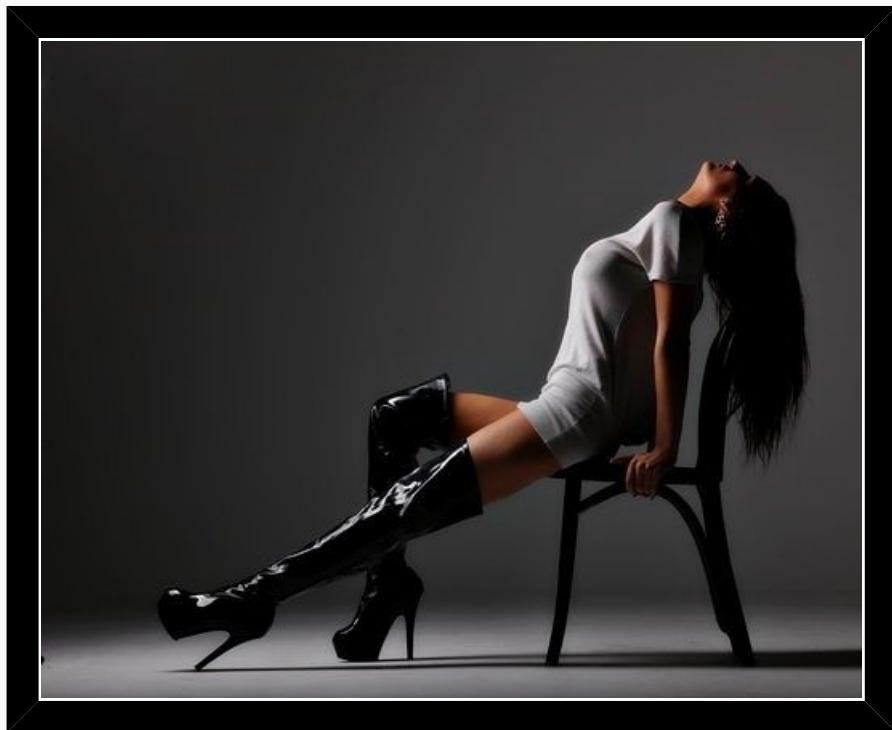
([http://www.filmphotoacademy.com/blog/jm\\_1760.jpg](http://www.filmphotoacademy.com/blog/jm_1760.jpg))

Microsoft says: a woman jumping up to catch a frisbee

Microsoft tags: , water, top, girl, playing, beach, air, surfing, jumping, board, catch, riding, frisbee, plate, standing

Google tags: barechestedness, muscle, male, arm, leg, art model, sense, hand, human body,

In both cases while there are more keywords from Microsoft they are far less accurate than those from Google. I'm especially impressed that it can sport "modern dance" as a particular genre.



(<http://www.filmphotoacademy.com/blog/POTD-0060.jpg>)

Microsoft says: a woman holding a cell phone

Microsoft tags: ,table,standing,computer

Google tags: sitting, performance-art, footwear, guitarist, high-heeled-footwear, leg, photo-shoot, modern-dance, sense

Just one other instance of Google being pretty spot on and Microsoft being way off.



([http://www.filmphotoacademy.com/blog/BTS\\_00000000\\_04.jpg](http://www.filmphotoacademy.com/blog/BTS_00000000_04.jpg))

Microsoft says: a man is walking down the street talking on a cell phone

Microsoft tags: outdoor, building, sidewalk, street, phone, city, walking, side, standing, graffiti, doing, riding, park, board, trick, people, sign, air, group

Google tags: art

It is not all one sided though. For this image above Google came back with one word whereas Microsoft has a pretty good selection that was fairly good. The Microsoft caption was very close.

So where does that get us? You could probably make some kind of image tag by using a combination of the Microsoft caption and the Google keywords. But there would also be an awful lot of junk and superfluous text in there too.

However, I am going to load up the alt text on the images on my blog using this information and overtime I'll try and get a picture of whether this is good for either people finding my images or my more general SEO position.

What I won't be doing is using any of these in human readable position. So for Instagram and the like I will stick to manually deciding on hash tags.

## Does size matter?

Yes, to an extent and in particular for face recognition. Both systems return more tags and more accurate tags with different size images. I tried 128 pixels, 512 pixels and 3600 pixels (longest side). The 128 pixel versions were considerably less accurate than the 512 pixels. Between the 512 and 3600 there were slight differences, but not significant.

If you are doing this yourself just for the keywords then 512 pixel images would be fine and saves hugely on the data being transported. Remember the API's have to upload each image in order to analyse them.

Let us use this image as an example, fed to the AI systems at three widths.



([http://www.filmphotoacademy.com/blog/Atest\\_5\\_3600.jpg](http://www.filmphotoacademy.com/blog/Atest_5_3600.jpg))

Microsoft caption:

128 pixels: person standing in a room

512 pixels: a woman sitting on a rug

3600 pixels: a woman in a white dress posing for a picture

The caption definition has clearly become more accurate as pixel count has increased.

Microsoft keywords

At 128 pixels: front standing young playing woman room ball

At 512 pixels: indoor woman dress girl posing room front sitting wearing standing young table bed red rug living court man

At 3600 pixels: indoor woman dress posing wearing girl front standing room table young rug red hat bed

At 128 pixels it isn't very convincing at all. The differences between 512 and 3600 are barely significant , a slight change in the order and the introduction of a "hat" but losing the "court" and "man".

Google keywords

At 128 pixels: dance performing arts team sport entertainment sports performance art gown

At 512 pixels clothing lady beauty dress sitting fashion leg model art model

At 3600 pixels lady beauty performance art dress fashion leg model photo shoot

Googles results are much more mixed, with performing arts coming and going. Note the sports entries in the 128 list, but these are gone from the later sizes.

Face Recognition

However for face recognition things fared much more badly, with faces not being recognised by either system at the smaller sizes. For example, both these images found a face at 3600 pixels, but failed at 512 pixels. Clearly something for some further investigation.



([http://www.filmpphotoacademy.com/blog/Atest\\_5\\_3600.jpg](http://www.filmpphotoacademy.com/blog/Atest_5_3600.jpg))



([http://www.filmphotoacademy.com/blog/Atest\\_4\\_3600.jpg](http://www.filmphotoacademy.com/blog/Atest_4_3600.jpg))

## Technical Conclusion:

If I was doing this again I would firstly experiment further at getting the image file size optimal. Worst comes to worst you just send gigabytes of images to the AI systems. Neither Microsoft or Google actually have a size limit. It just offends my engineers sense of efficiency.

You could use the Microsoft captions as hidden captions, but they are very bland.

I think the google keywords are more accurate - I would use them as the principal keywords, then add as a lower priority the keywords from Microsoft.

I would combine both systems for face detection. Google found 25% more faces than Microsoft but both sides had plenty of gaps where one system found a face and the other didn't.

Including such an automated system into a web based forum or photo display site is emminently practical. Both sites have plenty of sample code (none of which was any use to me mind since I wanted a VBA or classic ASP solution!)

Neither system is perfect, while either or both can do a lot of the leg work, you still need a human intervention on the edge cases.

A sample of around 100 images that were used in this test, along with the results returned can be downloaded here.

[http://filmphotoacademy.com/download/ai\\_vision\\_by\\_filmphotoacademy.pdf](http://filmphotoacademy.com/download/ai_vision_by_filmphotoacademy.pdf)  
([http://filmphotoacademy.com/download/ai\\_vision\\_by\\_filmphotoacademy.pdf](http://filmphotoacademy.com/download/ai_vision_by_filmphotoacademy.pdf))

I have added a link so you can download the code and try it yourself.

This code was put into a Microsoft Access database. It will probably run in Microsoft Word with minimal or no changes

You will need to go to Google Vision and Microsoft Vision and sign up to the API's to get your own keys.

Have fun - and I'd love to hear about what else you've done with this code.

[http://www.FilmPhotoAcademy.com/download/ai\\_vision\\_code.txt](http://www.FilmPhotoAcademy.com/download/ai_vision_code.txt)  
([http://www.filmphotoacademy.com/download/ai\\_vision\\_code.txt](http://www.filmphotoacademy.com/download/ai_vision_code.txt))

share: [email](mailto:?subject=Interesting%20Blog%20Page%20at%20FilmPhotoAcademy.com&body=This%20blog%20page%20is%20interesting.%20http://www.FilmPhotoAcademy.com/blog/default.asp?page=post&recordID=3499) ([mailto:?subject=Interesting Blog Page at FilmPhotoAcademy.com&body=This blog page is interesting. http://www.FilmPhotoAcademy.com/blog/default.asp?page=post&recordID=3499](mailto:?subject=Interesting%20Blog%20Page%20at%20FilmPhotoAcademy.com&body=This%20blog%20page%20is%20interesting.%20http://www.FilmPhotoAcademy.com/blog/default.asp?page=post&recordID=3499))

[facebook](https://www.facebook.com/sharer/sharer.php?u=http://www.FilmPhotoAcademy.com/blog/default.asp?page=post&recordID=3499) (<https://www.facebook.com/sharer/sharer.php?u=http://www.FilmPhotoAcademy.com/blog/default.asp?page=post&recordID=3499>) [twitter](https://twitter.com/home?status=http://www.FilmPhotoAcademy.com/blog/default.asp?page=post&recordID=3499) (<https://twitter.com/home?status=http://www.FilmPhotoAcademy.com/blog/default.asp?page=post&recordID=3499>)

[reddit](http://www.reddit.com/submit?url=http://www.FilmPhotoAcademy.com/blog/default.asp?page=post&recordID=3499) (<http://www.reddit.com/submit?url=http://www.FilmPhotoAcademy.com/blog/default.asp?page=post&recordID=3499>) [reddit img](http://www.reddit.com/submit?image=http://www.filmphotoacademy.com/blog/uploads/512/X4.jpg) (<http://www.reddit.com/submit?image=http://www.filmphotoacademy.com/blog/uploads/512/X4.jpg>)



(<http://www.usave.photography/>)

(<http://www.usave.photography/>) Would you like to take better photographs?

Would you like to take a big step forward in a photography career or as a keen enthusiast?

"Master Your DSLR" is a comprehensive course provided completely online, in your own time, at your own pace.

More Info (<http://http://ulearn.photography/>)

[more...](#) (<http://www.usave.photography/>)

← PREVIOUS (<http://www.filmphotoacademy.com/blog/model-of-the-day-01122016>)

→ NEXT

(<http://www.filmphotoacademy.com/blog/>)

## Related posts

Homage to Bill Brandt

(<http://www.filmphotoacademy.com/blog/default.asp?page=post&RecordID=705>)



(<http://www.filmphotoacademy.com/blog/default.asp?page=post&RecordID=705>)

AOP part 28: Keeping the Feel

(<http://www.filmphotoacademy.com/blog/default.asp?page=post&RecordID=763>)



(<http://www.filmphotoacademy.com/blog/default.asp?page=post&RecordID=763>)

AOP part 3: Strength

(<http://www.filmphotoacademy.com/blog/default.asp?page=post&RecordID=746>)



(<http://www.filmphotoacademy.com/blog/default.asp?page=post&RecordID=746>)

JammyLou Nude Photoshoot

(<http://www.filmphotoacademy.com/blog/default.asp?page=post&RecordID=393>)



(<http://www.filmphotoacademy.com/blog/default.asp?page=post&RecordID=393>)

## Categories

General Photography (<http://www.filmphotoacademy.com/blog/default.asp?category=8>)

