The New New Math: using sentiment analysis of mathematics word problems to gauge children's reactions to teaching 8-bit floating point arithmetic for the new California public school math curriculum

Lee "just a very concerned parent" Butterman
leebutterman@gmail.com

Abstract

California's new mathematics curriculum plans to replace calculus with data science. New GPUs used in data science/machine learning perform 32 quadrillion arithmetic operations a second, at 8-bit floating point precision. Students should know how to harness this power, with fundamentals like $10 \times 10 = 96$, and $16 + 1 = 16$, so we introduce a multiplication table for FP8. We can use sentiment analysis from large language models to compare negative/positive sentiments around 'ten times ten makes ninety six' versus 'ten times ten makes one hundred', and we find numerologically significant patterns in the results. Further, sentiment analysis indicates that 'ten times ten makes ninety six' is a much more positive sentiment than '10 x 10 = 96', corroborating our national fear of Arabic numerals and our large-scale adoption of word problems.

1 New New Math: Replacing calculus with data science

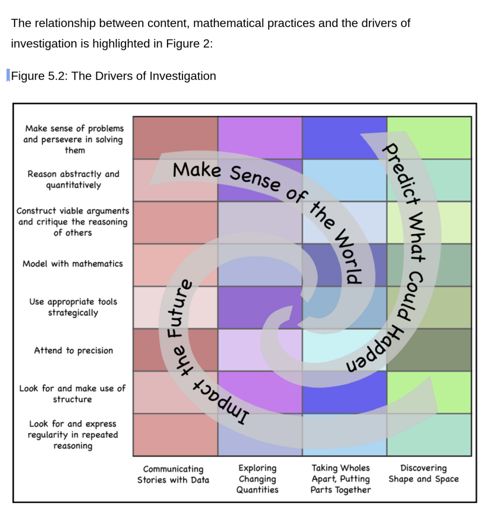


Figure 1: California new math guidelines, figure 5.2, articulates attending to precision.

California's guidelines [Commission, 2022] for its newest incarnation of New Math [Lehrer, 1965] are exciting. Instead of a child learning outmoded arcana like "the slope of a curve" or "piecewise differentiable functions", the child of tomorrow will learn new fresh relevant skills like *autograd* and `torch.nn.ReLU`.

SIGBOVIK 2022 (co-located at super-spreader event COVID19.BA.2.20220408.Allegheny), Pittsburgh, PA.

(a) 10.2 kW max, 32 petaFLOPS FP8, 0.24 petaFLOPS FP64, 8 rack units

(b) 8.88 kW max, 0 petaFLOPS FP8, 0 petaFLOPS FP64, 6+ seats

Figure 2: Equivalent energy usage for the affluent prosumer, different computation abilities

The guidelines about data science, in Chapter 5, emphasize keeping up to date with software and hardware advances: "Familiarity with technology and modern tools should progress through the grades." One long-useful tool [Krizhevsky et al., 2012] for data science/machine learning is the GPU.

Current GPUs are exceptionally powerful. The new H100 gpu [NVIDIA, 2022] can execute 0.03 quadrillion 64-bit floating point operations per second, and 4 quadrillion 8-bit floating point operations per second, and comes in a desktop form factor with 8 GPUs that uses more power than the largest Jacuzzi model [Jacuzzi, 2022] available.

Modern programmers are sufficiently surprised at the lack of precision of FP64 computation [Wiffin, 2022] in practice that we expect children to encounter such surprises in their schooling. The new H100 can execute over a hundred times as many FP8 operations per second as FP64, and FP8 is significantly less precise than FP64, trying to approximate all real numbers with only 256 numbers instead of 16 billion billion numbers.

2 A gentle foray into floating-point math

There are infinitely many real numbers, and some real numbers require infinite precision, and computers only have limited space. Some common representations of numbers are familiar, like an 8-bit unsigned integer representation ('uint8'), where the dynamic range of the representation is $0, 1, \ldots, 254, 255$.

FP64, 64-bit floating point, "double-precision" floating point, is well-established, since 1984, and is a slightly more complicated representation. Floating point representations are determined by (sign width, exponent width, mantissa width, exponent bias), where exponent and mantissa are unsigned positive integers from 0 to $2^{width-1}$. Specifically, a finite floating point number with sign/exponent/mantissa (S, E, M) and implicit bias B is interpreted as

$$-1^S \times 2^{E-B} \times \text{sign}(E).M$$

where sign is the signum function, twice the Heaviside step function minus 1. When E=$2^{width-1}$, the maximum exponent, and $M = 0$, that number represents $-1^S \times \infty$, $\pm\infty$. When E=$2^{width-1}$ and $M > 0$, that number represents a Not A Number number.[Plato, 384 BCE]

The FP64 representation is a (1-bit sign, 11-bit exponent, 52-bit explicit mantissa, -1023 bias) floating point representation, and can represent every integer from -2^{53} to 2^{53} (the $\text{sign}(E).M$ is a 1+52, or 53 bit number). The FP8 representation is a (1-bit sign, 5-bit exponent, 2-bit explicit mantissa, -15 bias) number, and between one and twenty can only represent 1, 1.25, 1.5, 1.75, 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8, 10, 12, 14, 16, and 20.

Note that FP8 only has 256 different real numbers to express all values from $-\infty$ to $+\infty$. With this paucity of choice, 3.5×3.5 ends up as 12, $3 \times 3 = 8$, $16 + 1 = 16$, and so on. This math is particularly useful in computation of weights of a neural network, where speed wins over accuracy often, and we can compute a hundred times as much arithmetic according to the new modern rules of $16 + 1 = 16$ than according to the antediluvian rules of $16 + 1 = 17$. We think the child of tomorrow will be at an advantage knowing another multiplication table.

FP64×	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	6	7	8	9	10
2	2	4	6	8	10	12	14	16	18	20
3	3	6	9	12	15	18	21	24	27	30
4	4	8	12	16	20	24	28	32	36	40
5	5	10	15	20	25	30	35	40	45	50
6	6	12	18	24	30	36	42	48	54	60
7	7	14	21	28	35	42	49	56	63	70
8	8	16	24	32	40	48	56	64	72	80
9	9	18	27	36	45	54	63	72	81	90
10	10	20	30	40	50	60	70	80	90	100

Figure 3: 1-10 multiplication table, with FP64 precision.

FP8×	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	6	7	8	8	10
2	2	4	6	8	10	12	14	16	16	20
3	3	6	8	12	16	16	20	24	28	32
4	4	8	12	16	20	24	28	32	32	40
5	5	10	16	20	24	32	32	40	48	48
6	6	12	16	24	32	32	40	48	56	64
7	7	14	20	28	32	40	48	56	64	64
8	8	16	24	32	40	48	56	64	64	80
9	8	16	28	32	48	56	64	64	80	96
10	10	20	32	40	48	64	64	80	96	96

Figure 4: 1-10 multiplication table, with FP8 precision.

3 Times tables, 1 to 10, at 8-bit and 64-bit floating point precision

This FP8 multiplication is refreshingly new, and thrillingly different from the multiplication table of a hundred years ago. There are only twenty two numbers to remember, and zero prime numbers lurking stealthily between the larger numbers in the multiplication table. We now turn to study how this new FP8 precision is emotionally received.

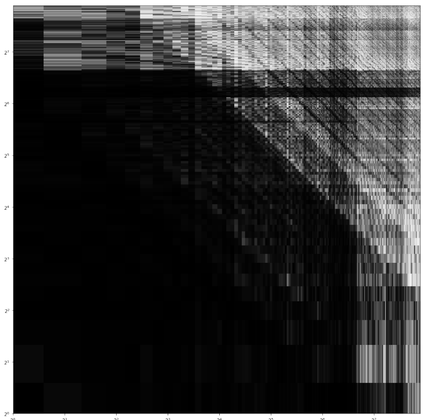
4 Understanding curriculum change: comparative sentiment analysis of math equations

Before introducing such a large change to the curriculum, it would be irresponsibly unfair not to know definitively how these changes impact the educational experience. Therefore, we perform an expensively thorough sentiment analysis on the resulting math equations, in several dimensions.

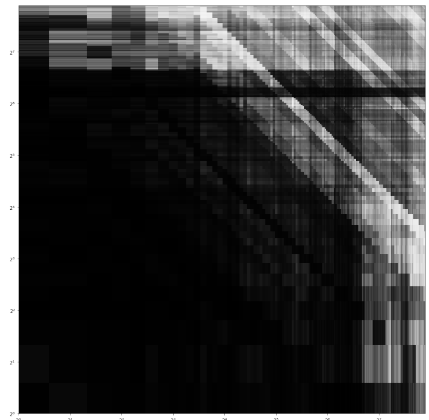
On a 1-to-240 square grid, we perform sentiment analysis for FP8 multiplication in English ('ten times ten makes ninety six') versus FP64 multiplication in English ('ten times ten makes one hundred'). We also perform sentiment analysis for FP8 multiplication in Arabic numerals ('$10 \times 10 = 96$') versus FP64 in Arabic numerals ('$10 \times 10 = 100$'). We also perform zero-shot question answering with masked language models, providing completions to masks like *Background: ten times ten makes ninety six. / Q: Is math fun? / A: <mask>*. This sentiment analysis is using distilbert-base-uncased-finetuned-sst-2-english, via a Huggingface pipeline. This masked language modeling is using distilroberta-base, also via a Huggingface pipeline.

Our results support the numerological emotional significance of certain numbers. When plotted on log-log axes, significantly positive or negative emotions jump out as black or white diagonal stripes respectively. We encourage teachers to treat numbers that multiply to these products with all the appropriate care when introducing them in class, particularly to younger children.

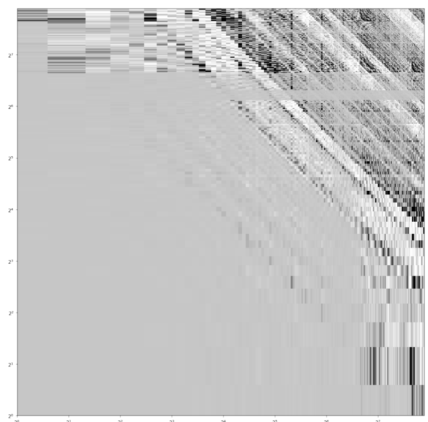
4.1 FP8 vs FP64 English sentiment comparison: ‘ten times ten makes ninety six’ versus ‘ten times ten makes one hundred’



(a) FP64: sentiments at (10,9) are for ‘nine times ten makes ninety’



(b) FP8: sentiments at (10,9) are for ‘nine times ten makes ninety six’



(c) FP8/FP64 sentiments, clipped at 0.5 (white) to 2 (black)

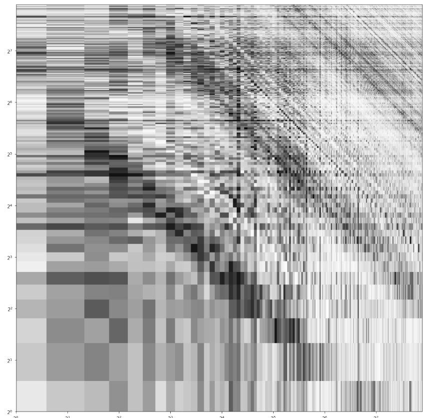
Figure 5: Sentiment analysis of multiplication tables from 1 to 240, at FP8 and FP64 precision, in plain English, individual and comparative. White is the minimum, and black is the maximum. Logarithmic axes. Note the light/dark diagonal bands that indicate aversion/fondness towards pairs of numbers that share a product.

As expected in Figure 5, the comparative sentiments are close (grey) for smaller numbers, because of the greater accuracy of floating point representations closer to zero. Note how the individual sentiments are both markedly less positive past the 2^{10} diagonal, indicating a deep human aversion to large numbers. Note the clear white stripes at FP8 for products of 2^{11} and 2^{13}, among others, showing the power of the stealthy large prime number.

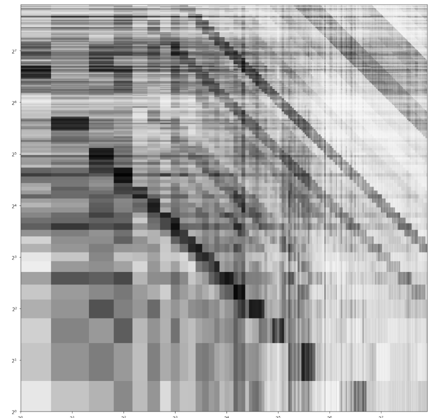
Also note that many rules of arithmetic do not apply to FP8 calculations, like associativity over addition. Similarly, sentiment is not commutative over multiplication. The sentiment of ‘twenty five times seventy five makes one thousand seven hundred ninety two’ is only 81.96% positive, while the sentiment of ‘seventy five times twenty five makes one thousand seven hundred ninety two’ is 94.22% positive. That increase of positivity is the dark band at the top of both images.

Instructors are encouraged to use numbers in this band as Preferred Multipliers for the multiplicands of their choice. Historically, mathematics instruction has treated multiplication as commutative, with minimal consideration towards directing numbers to the multiplier versus the multiplicand. Sentiment analysis shows that pupils may react differently to different rearrangements of equations, and may achieve higher test scores and may be more engaged when encountering a Preferred Multiplier.

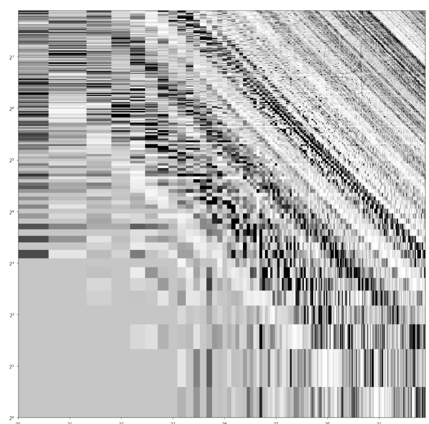
4.2 FP8 vs FP64 Arabic numeral sentiment comparison: ‘10×10=96’ versus ‘10×10=100’



(a) FP64: sentiments at (10,9) are for ‘9 x 10 = 90’



(b) FP8: sentiments at (10,9) are for ‘9 x 10 = 96’



(c) FP8/FP64 sentiments, clipped at 0.5 (white) to 2 (black)

Figure 6: Sentiment analysis of multiplication tables from 1 to 240, at FP8 and FP64 precision, in Arabic numerals, individual and comparative. White is the minimum, and black is the maximum. Logarithmic axes. Note the consistent light/dark diagonal bands that indicate aversion/fondness towards pairs of numbers that share a product.

Similarly to Figure 5, in Figure 6 the comparative sentiments are close (grey) for smaller numbers, because of the greater accuracy of floating point representations closer to zero, and note further that the grey square at the lower right is much smaller, indicating a much more perceptible difference to the sentiment of numbers in number problems compared to words in word problems.

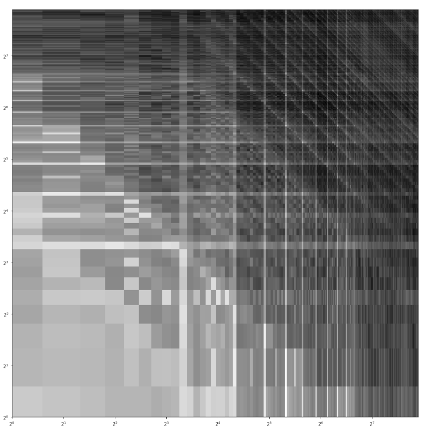
Note also how much lighter the individual sentiments are: this means that there is much more negative sentiment towards numbers than towards letters. This aligns to the innumerate bigoted xenophobia [Akyol, 2019] that causes Americans to poll negatively when asked about Arabic numerals.

Note the black stripe in the FP8 around products like 80 and 96. The sentiment of ‘6 x 16 = 96’ is 90.06% positive, whereas the sentiment of ‘8 x 16 = 128’ is 72.85%. Note further the FP8 stripes at products of 2^{11} and beyond.

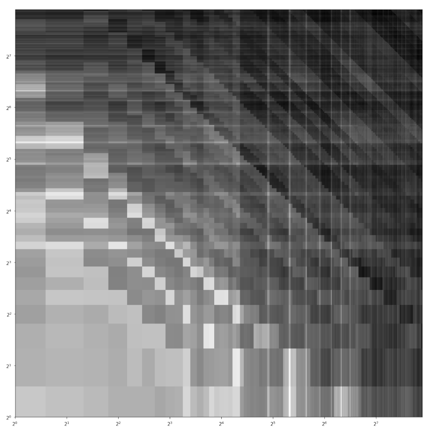
Further note that there is a much fainter and much larger band for Preferred Multipliers for FP8 (and even FP64), from just over 2^6 to just over 2^7, and that there is a faint band of Preferred Multiplicands from ≈35 to ≈50. Instructors are urged to not carelessly reuse Preferred Multipliers between numerical lessons and the surrounding discussion, and to not carelessly choose Preferred Multiplicands, and to understand how their choices may impact their pupils’ weekly standardized test scores.

4.3 FP8 vs FP64: do you like math? Math as fun+unpleasant horror movie

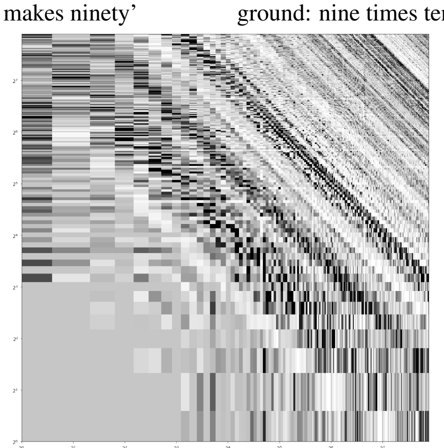
The next careful and thoughtful investigation is the relative fun of mathematics, while using FP8 precision versus FP64. For our survey, we computed the most likely ways to fill a mask in the following template string: *Background: five times five makes twenty four¶¶Q: Do you like math?¶A:* <*mask*>, substituting appropriate numbers for the multiplier, multiplicand, and the product. The masks filled only in a positive or negative assertion (these are high quality Large Language Models [Bender et al., 2021] after all). The value at each product is 0.5 $\pm$ the highest probability token, positive if positive, negative if negative.



(a) FP64: sentiments at (10,9) are for 'Background: nine times ten makes ninety'



(b) FP8: sentiments at (10,9) are for 'Background: nine times ten makes ninety six'



(c) FP8/FP64 sentiments, clipped at 0.5 (white) to 2 (black)

Figure 7: Mask fills for survey data from 1 to 240, at FP8 and FP64 precision, in English letters, individual and comparative. White is the minimum, and black is the maximum. Logarithmic axes. Note the consistent light/dark diagonal bands that indicate how miserable/enjoyable that product of numbers colors the attitude toward mathematics.

Note that 10 is what we refer to as both a Joyless Multiplier and Joyless Multiplicand, because its row and column are much lighter (less fun) than their surrounding values.

These results imply that mathematical statements have negative sentiment, and positive fun, and are more fun the larger they are, even as they are negative sentiments as the multiplier and multiplicand grow, like horror movies with entertainingly enormous monsters, or gargantuan public speaking events, or other such formative experiences. This formativity aligns with our idea of a mathematics education.

5 Conclusion

New hardware-accelerated numerical representations will shape how we teach children mathematics, data science, machine learning, and more. The newest, FP8, is on a GPU that can compute 32 million billion arithmetic operations per second with numbers in FP8 format. We can use sentiment analysis to show that mathematics according to FP8 rules is not strongly different in emotional state than according to FP64 rules. Sentiment analysis also shows how much we prefer numbers spelled out, which suggests new life-long learning opportunities. Masked language modeling shows that we find mathematics fun, and a promising future direction would be to compute Net Promoter Scores per mathematical statement with these large language models, to keep a finger on the pulse of the youth.

References

Mustafa Akyol. Who's Afraid of Arabic Numerals? 2019. URL https://www.nytimes.com/2019/06/04/opinion/arabic-numerals.html.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, page 610–623, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383097. doi: 10.1145/3442188.3445922. URL https://doi.org/10.1145/3442188.3445922.

Instructional Quality Commission. *2022 Revision of the Mathematics Framework*. 2022. URL https://www.cde.ca.gov/ci/ma/cf/.

Jacuzzi. J-495 Spacious Designer Entertainer's Hot Tub. 2022. URL https://www.jacuzzi.com/en-us/j-495-spacious-designer-entertainers-hot-tub/J-495.html.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *NeurIPS*, 2012.

Tom Lehrer. New Math. 1965.

NVIDIA. H100 Tensor Core GPU. 2022. URL https://www.nvidia.com/en-us/data-center/h100/.

Plato. *Euthydemus*. 384 BCE.

Erik Wiffin. 0.30000000000000004. 2022. URL https://0.30000000000000004.com/.