

intel.com/ai
#intelAI

AI: UNLEASHING THE
NEXT WAVE

 Artificial Intelligence

Решения Intel для искусственного интеллекта и анализа больших данных

Сергей Гололобов

НГУ/Intelligent
Computing

AI = ИИ (Искусственный Интеллект)

«Тот, кто станет лидером в этой области, будет властелином мира»

В.Путин, сентябрь 2017

Поток данных нарастает...

К 2020...



Пользователь интернета будет генерировать
~1.5 ГБ трафика в день



Больница будет генерировать
3,000 ГБ в день



Беспилотный автомобиль будет генерировать
4,000 ГБ в день



Авиалайнер будет генерировать
40,000 ГБ в день



Завод будет генерировать
1,000,000 ГБ в день



Радар	~10-100 КБ	в секунду
Сонар	~10-100 КБ	в секунду
GPS	~50 КБ	в секунду
Лидар	~10-70 МБ	в секунду
Камеры	~20-40 МБ	в секунду



All numbers are approximated
<http://www.cisco.com/c/en/us/solutions/service-provider/vni-network-traffic-forecast/infographic.html>
http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud_Index_White_Paper.html
<https://datafloq.com/read/self-driving-cars-create-2-petabytes-data-annually/172>
http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud_Index_White_Paper.html
http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud_Index_White_Paper.html



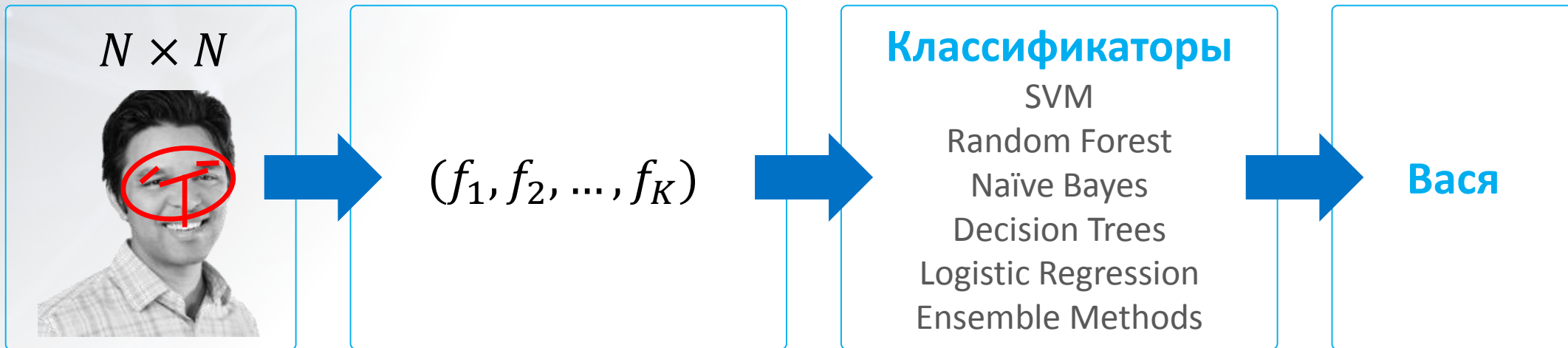
Что такое искусственный интеллект?

Искусственный интеллект

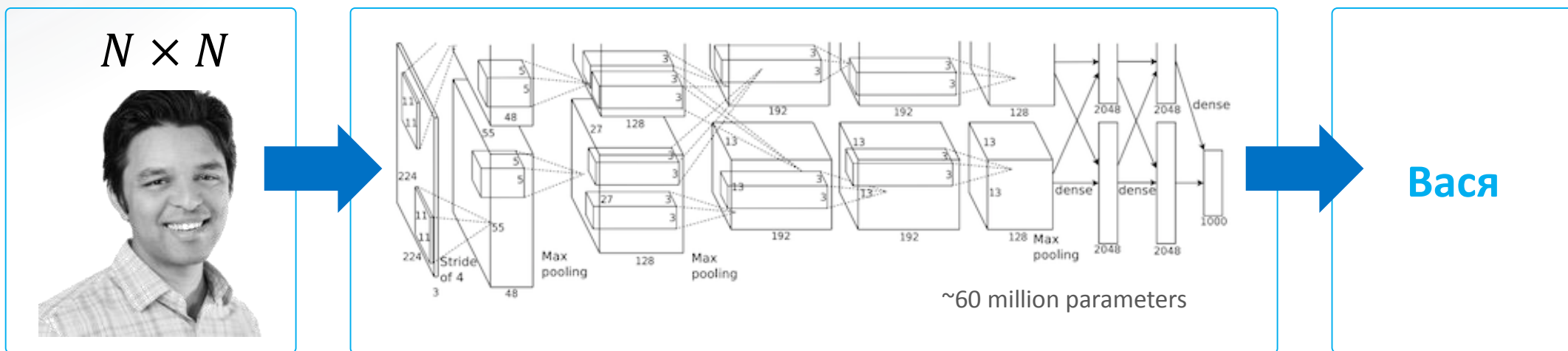


Машинное обучение

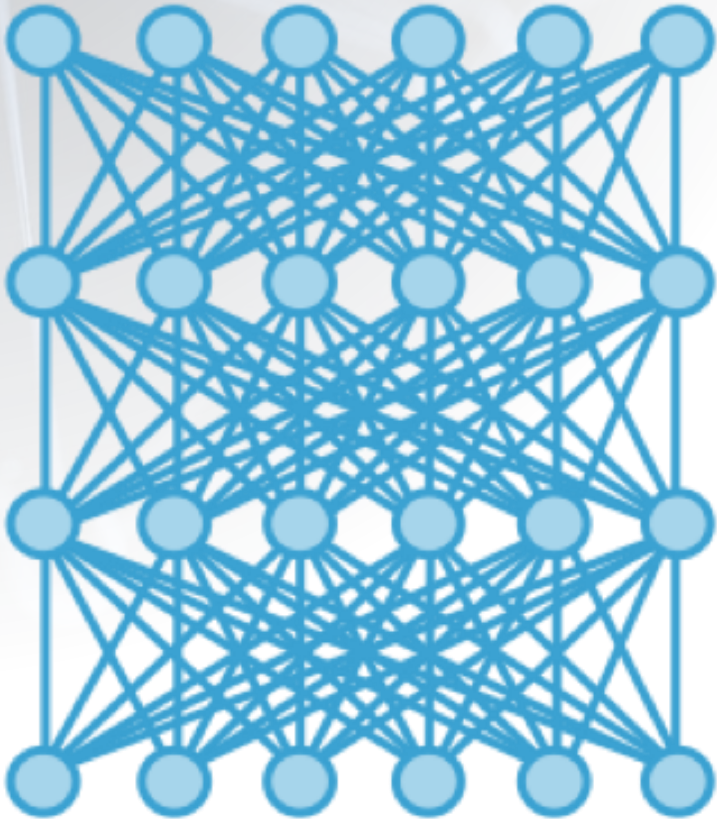
Классический
ПОДХОД



Глубокое
обучение



Глубокое машинное обучение



Обнаружение свойств в данных

Извлечение свойств на разных уровнях абстракций

Результат улучшается при увеличении количества данных

Способность работать со сложными данными

Искусственный интеллект для аналитики



AI

является огромной
самодостаточной
областью и
стратегическим
инструментом в
деле развития
аналитики

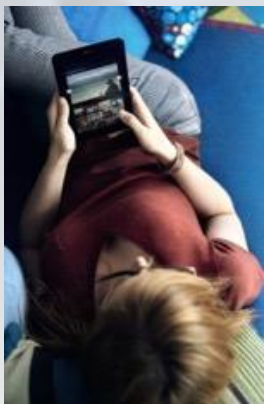
Систематизация нашего подхода



Подходы

Варианты

AI меняет окружающий мир



Клиент

Медицина

Финансы

Торговля

Управление

Энергетика

Транспорт

Промышленность

А также

Помощники
Чат-боты
Поиск
Персонализация
Виртуальная реальность
Роботы

Диагностика
Разработка лекарств
Лечение
Исследования
БАДы

Торговля
Поиск махинаций
Исследования
Личные финансы
Управление рисками

Поддержка
Опыт
Маркетинг
Продажи
Лояльность
Поставки
Безопасность

Оборона
Разведка
Безопасность
Гражданское общество
Умные города

Поиск нефти и газа
Умные решения
Сохранение природы

Беспилотные автомобили
Автоматическая доставка грузов
Авиа и космическая промышленность
Служба спасения

Автоматизация производства
Умная поддержка
Агротехника

Реклама
Образование
Игры
Службы
Связь
Спорт

Source: Intel forecast

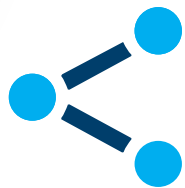
Примеры AI

Различные подходы в ИИ к различным задачам



Глубокое машинное обучение

Распознавание образов, речи, шаблонов и т.п.



Классическое машинное обучение

Статистические задачи, рекомендации и т.п.



Интеллектуальные системы

Многофакторный анализ, поиск мошенничеств, анализ клиентов и т.п.

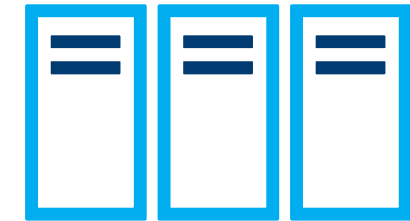
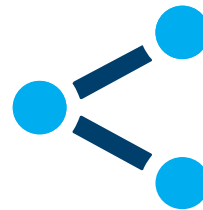


Будущее

Выравнивание последовательностей в биологии, бинарные нейронные сети и т.п.

Пример комплексного решения

Беспилотный автомобиль





Искусственный интеллект с Интел

Решения Интел для ИИ



Дизайн

Глубокое машинное обучение и не только

Технология Intel® Nervana™ разработана специально для нейронных сетей и с учётом их специфики. И дальнейшие улучшения разрабатываются прямо сейчас.



Разработка

Новые возможности для ИИ

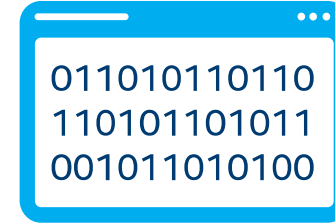
Технология Intel® Nervana™ открывает простор для получения новых результатов в области ИИ



Масштабирование

От оконечных устройств до ЦОДов

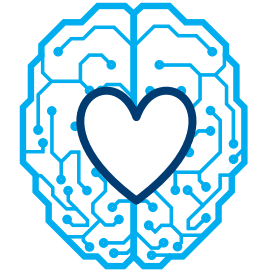
Интел имеет все компоненты, которые позволят всем желающим создать эффективные решения от А до Я.



Обработка данных

Новые приложения

Интел расширяет вычислительные возможности для того, чтобы все желающие могли получать новые результаты из собранных данных с помощью продуктов, созданных специально для задач ИИ.



Польза для людей

Максимум пользы
Интел инвестирует в приложения, которые приносят пользу обществу. Интел сотрудничает с правительством, бизнесом и активными гражданами для ускорения процесса преобразований.

Портфолио Intel® Nervana™

Создано для

ИИ

Практика



Платформы

Intel® Nervana™ Cloud & Appliance
Intel® Nervana™ DL Studio

Intel® Computer
Vision SDK

Movidius
(VPU)



Инструменты



Mlib
BigDL



mxnet



torch

Caffe

Caffe2



theano

Библиотеки



Intel Python
Distribution

Intel® Data Analytics Acceleration
Library
(DAAL)

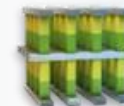
Intel® Nervana™ Graph*
Intel® Math Kernel Library
(MKL, MKL-DNN)

Оборудование

Вычисления



Системы хранения



Сети



*В планах
Other names and brands may be claimed as the property of others.

Искусственный Интеллект от А до Я



ЦОДы

Обработка больших/поточковых данных

Проводные и
беспроводные каналы
связи



ШЛЮЗЫ

Пересылка данных между
устройствами

- Протоколы передачи данных
- ✓ Безопасные
 - ✓ С высокой пропускной способностью
 - ✓ В режиме реального времени



Оконечные устройства

Устройства с низким энергопотреблением,
адаптированные к нуждам пользователя



Процессоры Intel® Xeon®



Процессоры* Intel® Xeon Phi™



Процессоры Intel® Core™ & Atom™

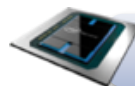
CPU+



Процессорная графика Intel®



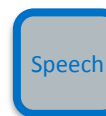
ПЛИСы Intel® (FPGA)



ИССН* Crest (ASIC)



ИССН Movidius VPU



Speech

ИССН* Intel® GNA (IP)

*Future

Инструменты

Выбирайте тот инструмент, который Вам нравится



Caffe



Caffe2



Microsoft CNTK

theano

mxnet



BigDL MLlib

и ещё больше инструментов будет доступно через Intel® Nervana™ Graph

Справочный инструмент компании Интел, который ориентирован на максимальную производительность на всём наборе архитектур компании intelnervana.com/neon

✓ оптимизировано для архитектур компании Интел

О готовности решений смотрите информацию на сайте компании

Права на другие торговые марки могут принадлежать другим компаниям.

ПО Intel® Nervana™ Deep Learning Studio

Уменьшение времени на разработку для получения быстрых результатов

Обработка данных
Сортировка данных
Маркировка данных

Блочная классификация
Сжатие моделей
Выпуск моделей
Экспорт на оконечные устройства



Многопользовательский режим
Интерактивное взаимодействие
Образцы решений

Быстрое обучение
Блочное обучение
Отслеживание прогресса
Распределение по узлам
Аналитика и визуализация
Параметрическая оптимизация

Пока доступно только в ЦОД Intel® Nervana™ Cloud; с 4 квартала 2017 будет доступно всем

Платформа Intel® Nervana™

Самая продвинутая платформа в мире для глубокого машинного обучения

Полноценная, удобная система для разработки и выпуска высокоточных решений на базе искусственного интеллекта за малое время:



Уменьшение времени на разработку



Ускорение вычислений



Дополнительные бонусы



Images



Video



Text



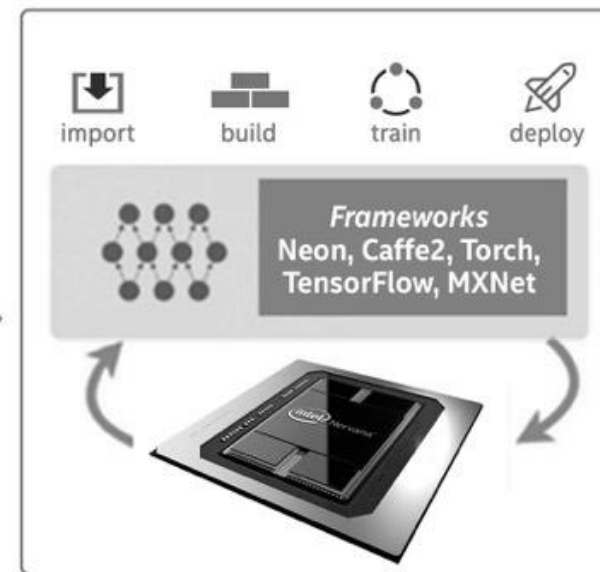
Speech



Tabular



Time Series



Available as hosted cloud services or on-prem in your datacenter



*Other names and brands may be claimed as the property of others.

ПО для шлюзов/оконечных устройств

Оптимизируйте модели глубокого машинного обучения для получения результатов на уровне шлюзов/оконечных устройств



У нас есть возможность **тонкой настройки** конечного результата с помощью инструментов Intel® MKL-DNN, c1DNN, OpenCL и OpenVX

Что мы делаем:

**Берём
обязательство**

Исполнительный директор лично руководит направлением ИИ, включая основные поглощения и формирование команды ИИ

**Выделяем
ресурсы**

включая полный набор инструментов для максимально полного использования технологий Искусственного Интеллекта

**Создаём
окружающий мир**

большая часть задач с использованием искусственного интеллекта решается с помощью архитектур компании Интел

**Предлагаем
технологии**

построенные на основе самых плотных и энергоэффективных подходов к созданию транзисторов на планете

Используем опыт

нескольких успешных этапов эволюции компьютерных вычислений



Мы обращаемся к ВАМ!

Начните работу с ИИ сегодня

Учитесь

Исследуйте

Включайтесь

Информация находится на сайте
www.intelnervana.com

Используйте оптимизированные
инструменты компании Интел

Свяжитесь с представителями
компании для получения
дополнительной информации о
существующих возможностях



Legal Disclaimer & Optimization Notice

- INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS”. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO THIS INFORMATION INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.
- Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.
- Copyright © 2016, Intel Corporation. All rights reserved. Intel, Pentium, Xeon, Xeon Phi, Core, VTune, Cilk, and the Intel logo are trademarks of Intel Corporation in the U.S. and other countries.

Optimization Notice

Intel’s compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

Notice revision #20110804

Backup



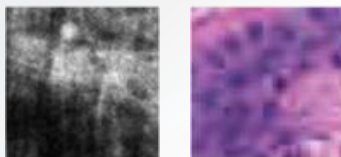
Глубокое машинное обучение в жизни

Медицина: поиск опухолей

Ткань

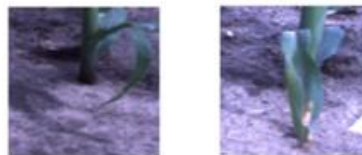


Опухоль



Сельское хозяйство

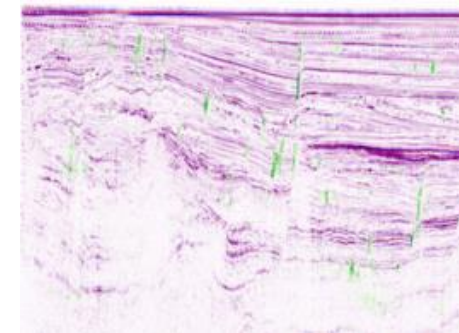
Растение



Сорняк



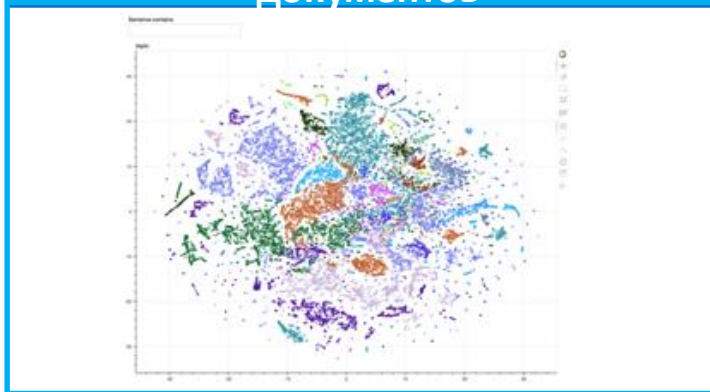
Добывающая промышленность



Управление голосом



Финансы: классификация документов



Анализ генных последовательностей

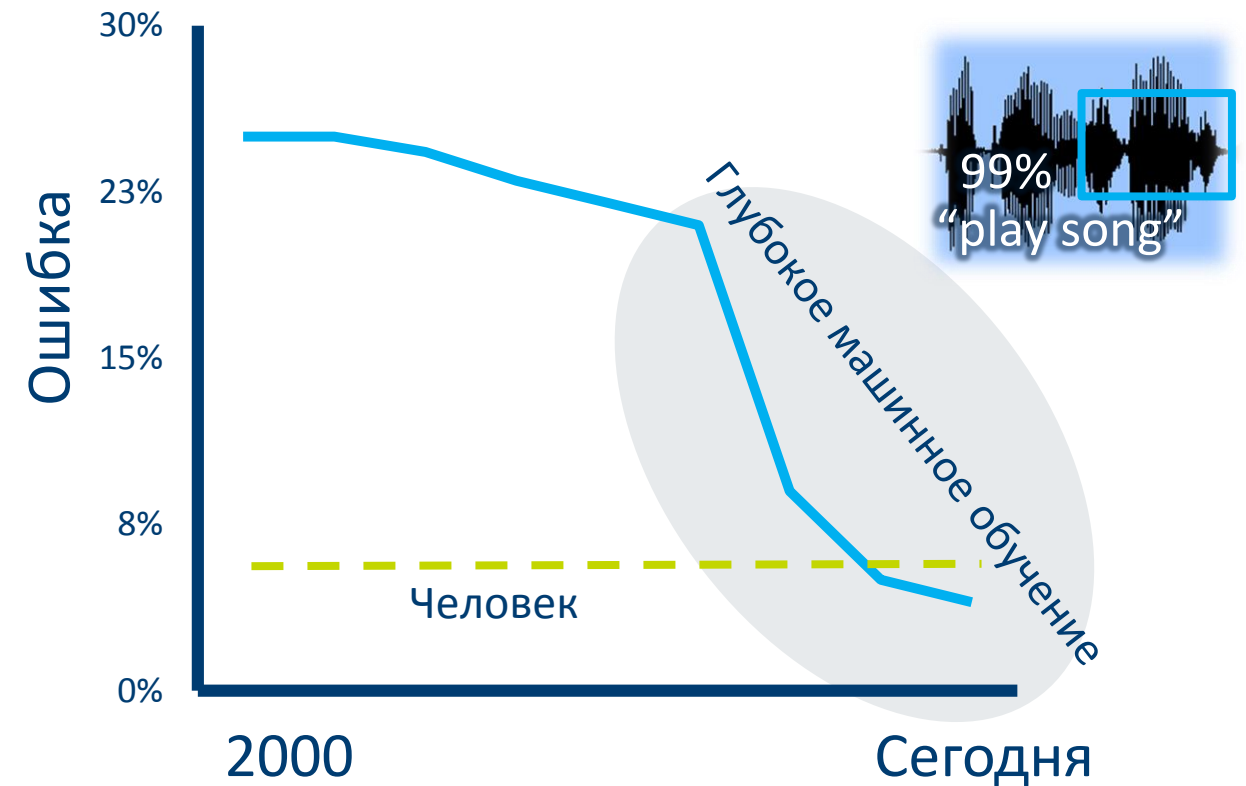


Достигнутые результаты

Распознавание образов



Распознавание речи

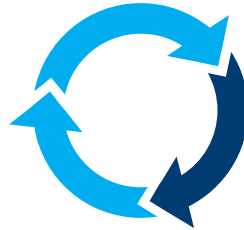


Intel® Xeon® processor scalable family

Scalable performance for widest variety of AI & other datacenter workloads – including deep learning



Most **agile** AI platform



Built-in ROI

Begin your AI journey today using existing, familiar infrastructure



Potent performance

Train in ~~days~~ HOURS with up to 113X² perf vs. prior gen (2.2x excluding optimized SW¹)



Production-ready

Robust support for full range of AI deployments

^{1,2}Configuration details on slide: 4, 5, 6
Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit: <http://www.intel.com/performance> Source: Intel measured as of November 2016
Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.
Notice Revision #20110804

ИИ



ЦОД

Многоцелевой



Процессоры Intel® Xeon®
Универсальная платформа

Scalable performance for widest variety of AI & other datacenter workloads – including deep learning training & inference

Параллельный



Intel® Xeon Phi™ Processor (Knights Mill[†])
Faster DL Training

Scalable performance optimized for even faster deep learning training and select highly-parallel datacenter workloads*

Гибкий

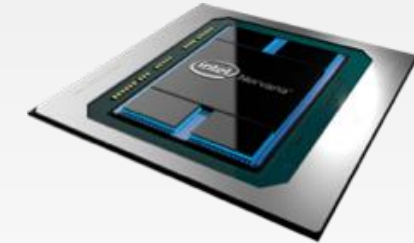


Intel®
FPGA

Enhanced DL Inference

Scalable acceleration for deep learning inference in real-time with higher efficiency, and wide range of workloads & configurations

Обучающий



Crest Family[†]

Deep learning by design

Scalable acceleration with best performance for intensive deep learning training & inference

[†]Условное название будущего продукта

Производительность сравнивается исключительно между продуктами компании Интел

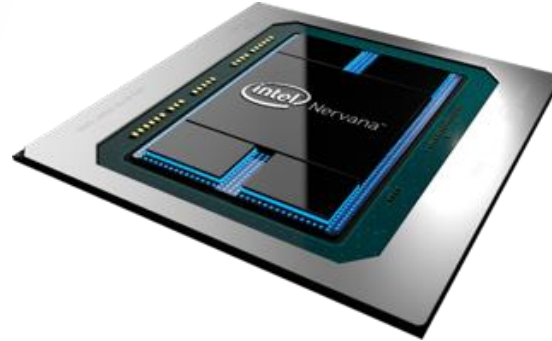
*Knights Mill (KNM); избранные = хорошо параллелизуемые приложения, работающие в одинарной точности, которые масштабируются на 100 и более потоков, векторизуются, и могут использовать высокую пропускную способность памяти, например, обратная миграция, машинное обучение и т.п.

Все продукты и системы, даты и целевые показатели являются предварительными и могут быть изменены без уведомления.

Crest family



Deep learning
By design



Scalable acceleration with best performance for intensive deep learning training & inference, period

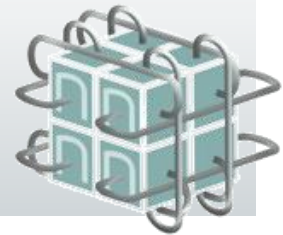
Custom hardware Blazing data access

- Unprecedented compute density
- Large reduction in time-to-train

- 32 GB of in package memory via HBM2 technology
- 8 Tera-bits/s of memory access speed

High-speed scalability

- 12 bi-directional high bandwidth links
- Seamless data transfer via interconnects



¹Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your actual performance. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit: <http://www.intel.com/performance> Source: Intel measured as of November 2016. Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Notice Revision #20110804

ИИ



шлюзы/оконечные устройства

all purpose

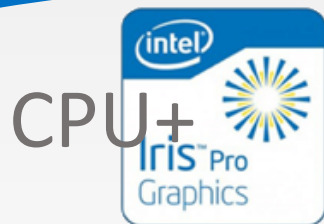


Intel® Processors

agile AI Platforms

Range of performance and power for widest variety of AI, gateway & edge workloads – including deep learning inference

graphics accel



Intel® Processor Graphics*

higher DL inference

Higher deep learning inference throughput and performance for graphics-intensive gateway & edge workloads**

Flexible accel



Intel® Arria® 10 FPGA

enhanced dl inference

Acceleration for deep learning inference in real-time with higher efficiency, and wide range of workloads & configurations

vision

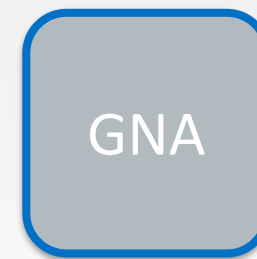


Movidius Myriad

computer vision

Ultra-low power computer vision engine using deep learning inference in gateway and devices

speech



Intel® GNA† (Future)

speech recognition

Ultra-low power speech recognition engine using deep learning inference in devices



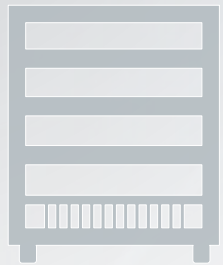
†GNA = Gaussian Mixture Model (GMM) and Neural Network Accelerator (IP block)

*Intel® HD Graphics, Intel® Iris™ Graphics, Intel® Iris™ Pro Graphics

**Graphics-intensive workloads include, for example, transcode, edge detection, image enhancement, 3D remote workstation, rendering, encoding, OpenCL/GL...

All products, computer systems, dates, and figures specified are preliminary based on current expectations, and are subject to change without notice.

Дальнейшее развитие



mainframes



Standard s-based servers



Cloud computing

- ☑ Data deluge
- ☑ COMPUTE
- ☑ breakthrough Innovation surge

Artificial intelligence

AI Compute Cycles will grow
by 2020 12X

Source: Intel forecast